

A reduced order Kalman Filter model for sequential Data Assimilation of turbulent flows

M. Meldi^{*a}, A. Poux^{**a}

^a*Institut PPRIME, Department of Fluid Flow, Heat Transfer and Combustion, CNRS - ENSMA - Université de Poitiers, UPR 3346, SP2MI - Teleport, 211 Bd. Marie et Pierre Curie, B.P. 30179 F86962 Futuroscope Chasseneuil Cedex, France*

Abstract

A Kalman filter based sequential estimator is presented in the present work. The estimator is integrated in the structure of segregated solvers for the analysis of incompressible flows. This technique provides an *augmented* flow state integrating available observation in the CFD model, naturally preserving a zero-divergence condition for the velocity field. Because of the prohibitive costs associated with a complete Kalman Filter application, two model reduction strategies have been proposed and assessed. These strategies dramatically reduce the increase in computational costs of the model, which can be quantified in an increase of 10% – 15% with respect to the classical numerical simulation. In addition, an extended analysis of the behavior of the numerical model covariance Q has been performed. The results have shown that optimized values are strongly linked to the truncation error of the discretization procedure. The estimator has been applied to the analysis of a number of test cases exhibiting increasing complexity, including turbulent flow configurations. The results show that the *augmented* flow successfully improves the prediction of the physical quantities investigated, even when the observation is provided in a limited region of the physical domain. In addition, the present work indicates that these Data Assimilation techniques, which are at an embryonic stage of development in CFD,

^{*}Corresponding author, marcello.meldi@ensma.fr

^{**}Current address : CNRS UMR 6614 CORIA - Université de Rouen - Site Universitaire du Madrillet, 675, avenue de l'Université, BP 12, 76801 Saint-Etienne du Rouvray Cedex, France

can be pushed even further using the *augmented* prediction as a powerful tool for the optimization of the free parameters in the numerical simulation.

Keywords: Kalman Filter, turbulent flows, Computational Fluid Dynamics

1. Introduction

The accurate prediction of turbulent flow configurations is one of the ultimate open challenges in fluid mechanics studies. Most of industrial / environmental applications aim to provide accurate and robust estimation of aspects which are governed by turbulence statistical moments such as aerodynamic forces, transport of particles and heat exchange. Traditional investigative tools, such as experiments and numerical simulation, are not completely successful in producing robust descriptions of turbulent configurations, because of important fundamental drawbacks.

Measurements obtained via Experimental Fluid Mechanics (EFD), such as those sampled by surface sensors, provide a local description of flow dynamics. Because of the non-linear, strongly inertial behavior of the flow, the determination of a complete map of flow behavior is problematic. Information can be reconstructed by the use of reduced-order models, such as POD [1]. However, these approximated models usually provide an incomplete reconstruction of turbulent flows. One of the reasons is that strong non-linear interactions occur between the modeling error and the bias associated to the measurement, which is tied to epistemic uncertainties in the experimental sampling. This aspect, which is amplified by the multi-scale nature of turbulence, usually results in poor characteristics of robustness and precision of the reduced-order model.

Under this perspective, the use of Computational Fluid Dynamics (CFD) can provide more complete maps of flow characteristics, including regions of the physical domain where experimental sampling is problematic because of structural difficulties. However, numerical simulation is affected as well by errors / epistemic uncertainties. In this case, the parametric set-up of the simulation (physical characterization of the flow, boundary conditions...) can not exactly

reproduce the subtle perturbations and in-homogeneity of the real flow, which are unknown a priori. This is particularly true for turbulence, where small perturbations present in the environment are amplified and they ultimately drive the evolution of the flow. In addition, owing to computational resources constraints, analyses of very high Reynolds flows are presently limited to reduced order simulations via RANS / LES approaches [2, 3, 4]. Turbulence / subgrid scale modeling is usually a main source of error in CFD, in particular because of its non-linear interaction observed with the numerical / boundary condition error. Thus, while both EFD and CFD are affected by bias, the confidence level of the results is affected by uncertainties of a completely different nature. This is the reason why the comparison of experiments and numerical results is a complex task even for the classical case of grid turbulence decay [5].

In the last decade, new methodological approaches coming from Estimation Theory (ET) have been employed by the fluid mechanics community to obtain an optimized prediction of flow configurations. ET is a branch of statistics dealing with the estimation of optimal parametric description, using data which is affected by a level of uncertainty / stochasticity [6]. In particular, Data Assimilation (DA) includes a wide spectrum of tools which aim at the estimation of an optimal state integrating a model and observations which are affected by uncertainties. They are usually referred to as *estimators*. Some studies in EFD have been proposed, where the DA tool combines experimental sampling with reduced order numerical solvers in order to provide the zero-divergence condition of incompressible flows [7, 8]. Early CFD applications mainly deal with variational approaches based on the adjoint method, which have been a classical choices for meteorological studies since the 1970s [9]. These methods are defined as an optimization problem where a given measure is minimized under the constraint of the governing equation. The resolution of this problem determines the distribution of a basis of parameters (typical choices are boundary / initial conditions) which optimizes the flow configuration. Recent applications deal both with fundamental studies [10, 11] and industrial oriented analyses [12, 13]. While these approaches are very precise and they allow for sensitivity analyses

considering a very large number of variables, their application to turbulent flow investigation over a long observation window is problematic [13, 14]. A much less commonly investigated path for CFD is represented by the use of sequential methods, which are based on Bayesian inference and they occasionally require the resolution of Riccati-type equations. Examples based on techniques such as the Kalman filter [15] or the ensemble Kalman filter [16] have been reported in the literature [17, 18, 19, 20, 21].

These techniques are showing enormous potential, because the coupling between experimental / numerical data can potentially exclude the bias which can not be identified in the two methods alone. However, reliable tools for the optimal prediction of complex flow configurations are still far out of reach, because of the level of maturity of application of these techniques in fluid mechanics analyses. In fact, these methods are still largely unexplored, in particular for the analysis of turbulent flow configurations. In the present work we propose a methodological approach (estimator) for sequential Data Assimilation, which efficiently integrates information (usually experimental data) in CFD solvers. The approach is based on a reduced order Kalman filter [15], which exploits structural characteristics of the segregated solvers commonly implemented in commercial CFD software. While similar approaches have been rigorously derived for coupled solvers for incompressible flows [22], the present model proposes practical, computational inexpensive solutions for the analysis of three dimensional turbulent flow configurations.

The paper is structured as follows. In Section 2 the state of the art is presented introducing all the background elements which synergically interact in the estimator. These elements include a description of the numerical solver as well as the Kalman Filter. In Section 3 the Kalman Filter based estimator is introduced and discussed. In Section 4 an extensive analysis of the structure of the model covariance matrix Q is performed. In section 5 the estimator is applied to the analysis of laminar flows. In particular, the property of synchronization of the model with available observation is investigated. In Section 6 the analysis is extended to turbulent flow configurations and their statistical behavior. In

Section 7 future development are discussed, including parametric optimization of the numerical model. Finally, in Section 8 conclusions are drawn.

2. Numerical & Methodological Ingredients: State of the Art

The basic elements for the development of the reduced-order Kalman filter estimator are now introduced. This includes a description of the segregated numerical CFD solver used for the resolution of the Navier-Stokes equations, as well as fundamental elements for Kalman Filter application.

2.1. The Pressure Implicit with Spitting of Operator (PISO) algorithm

Most of the CFD solvers available in commercial and open-source codes are based on numerical discretization of the Navier-Stokes equations. For incompressible flows of Newtonian fluids, the evolution of the velocity \mathbf{u} and the normalized pressure $p = p'/\rho$ is described as:

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{u} \mathbf{u}) - \nabla \cdot (\nu \nabla \mathbf{u}) = -\nabla p \quad (1)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (2)$$

$\nabla \cdot$ and ∇ are the divergence operator and the gradient operator, respectively. In the case of reduced-order simulation, simplified variables are considered (average velocity $\bar{\mathbf{u}}$ in the case of RANS, filtered velocity \mathbf{u}^* in the case of LES, ...) and a term T representing the turbulence / subgridscale modeling must be included in the vectorial equation 1.

Several different discretization strategies can be proposed for the numerical resolution of this system. In the following, the Pressure Implicit with Splitting of Operator (PISO) [23] as implemented in the open-source code *OpenFOAM* [24] is presented. This code, which elaborates a finite volume discretization of the Navier-Stokes equations, has been identified as the best demonstrator for this research activity thanks to the flexibility of its algorithmic structure and the ease in implementation.

At the beginning of every time step a *prediction* of the flow field is performed discretizing the momentum equation 1. In particular, an implicit discretization of the time derivative is applied. The PISO algorithm operates on the discretization at this level, simplifying the non-linear term in equation 1. This approximation, which relies on the hypothesis of small values of the time step Δt , resolves the non-linearity of the system via a linearization of the new state around the source state (previous time step). The approximation error of this operation is proportional to Δt^2 [25]. The resulting discretized system, known as *predictor*, is:

$$a_P \mathbf{u}_P = - \sum_N a_N \mathbf{u}_N + \Phi_0(\mathbf{u}_0) - \nabla p = \Phi(\mathbf{u}) - \nabla p \quad (3)$$

Here, the term with subscript 0 indicates a source velocity from previous time steps. The subscript P represents the discretized velocity field in the considered mesh element, while the subscript N indicates its neighbors. The coefficients a_P and a_N represent the result of the discretization process of the velocity budget terms in equation 1. As a first guess, the term ∇p is calculated using the pressure field at the previous time step. Because of the underlying hypothesis of the PISO algorithm over Δt , a Courant-Friedrick limit (CFL) for numerical stability must be satisfied for the time advancement. However, the resulting discretized system can be investigated by the resolution of a linear system. The coefficient a_P and the operator Φ capture the essential non-linearity of equation 1. Instead of performing a computationally expensive matrix inversion, the predictor is usually solved via iterative resolution techniques. In addition, the pressure-velocity coupling is avoided performing an iterative loop. First, it is observed that the velocity field predicted by equation 3 does not usually comply with the zero-divergence condition:

$$\nabla \cdot \mathbf{u} = \sum_f S \times \mathbf{u}_f = 0 \quad (4)$$

$$\mathbf{u}_f = \left(\frac{\Phi(\mathbf{u})}{a_P} \right)_f - \left(\frac{\nabla p}{a_P} \right)_f \quad (5)$$

where the subscript f stands for an interpolation on the face centers of the mesh element and S is the corresponding surface area. If equation 4 and 5 are combined, a Poisson equation for the pressure field is obtained:

$$\nabla \cdot \left(\frac{\nabla p}{a_P} \right) = \nabla \cdot \left(\frac{\Phi(\mathbf{u})}{a_P} \right) = \sum_f S \times \left(\frac{\Phi(\mathbf{u})}{a_P} \right)_f \quad (6)$$

The pressure field calculated in equation 6 does not necessarily satisfy equation 3. This value is used to correct the velocity field obtained in the *predictor* step. This operation is referred to as *corrector* step. The loop updating the pressure / velocity fields continues until a prescribed convergence criterion is satisfied. Summarizing, the loop can be structured in the following phases:

1. A first prediction of the velocity field is obtained by the discretized momentum equation 3. This velocity field is usually not divergence-free.
2. A Poisson equation is resolved to update the pressure field p as a function of the new predicted velocity field. The pressure gradient does not usually satisfy equation 3, so that the velocity field can be updated accounting for the new information.

In practice, the physical system bounces from a momentum-predicted state to a zero-divergence condition until the convergence is reached. An essential feature is that, while the Navier–Stokes equation are strongly non-linear, the discretization proposed includes the non-linearity in a linear system resolution, with an error scaling as Δt^2 . This last aspect, which is shared by a number of CFD numerical strategies, is essential for the integration of the proposed estimator, as it will be shown in Section 3.

2.2. The Kalman Filter

The Kalman Filter [15] is a powerful sequential tool used in Data Assimilation studies. It allows for robust prediction of an optimized physical state, accounting for the level of uncertainty in the prediction provided by multiple investigative tools. Let us consider a physical phenomenon \mathbf{u} described by the following discrete system at the instant k :

$$\mathbf{u}_k = \Phi_k \mathbf{u}_{k-1} + B_k \mathbf{c}_k + \mathbf{w}_k \quad (7)$$

$$\mathbf{z}_k = H_k \mathbf{u}_k + \mathbf{v}_k \quad (8)$$

Equation 7 represents a *model* for the time evolution of \mathbf{u} . A similar structure with respect to the *predictor* equation 3 is recognizable. On the other hand, the term \mathbf{c} represents a source term that can be manipulated by the user. Equation 8 represents available *observation* \mathbf{z} of the same physical phenomenon at the instant k . Both the equations are affected by a degree of epistemic uncertainty, represented by the terms $\mathbf{w}_k = \mathcal{N}(0, Q_k)$ and $\mathbf{v}_k = \mathcal{N}(0, R_k)$. This uncertainty can represent both bias associated with the prediction / observation as well as measurement noise. The terms w and v are supposed to be zero-mean and they are characterized by a time-dependent variance Q_k and R_k , respectively. The smaller the value of the variance, the higher the confidence in the model / observation.

The Kalman filter provides an optimized prediction of \mathbf{u} , which will be referred to as $\hat{\mathbf{u}}$, accounting for the level of confidence in the model and in the available observation. More precisely, the optimized state provides a minimization of the error covariance of the physical system $P = E((\mathbf{u} - \hat{\mathbf{u}}) \times (\mathbf{u} - \hat{\mathbf{u}})^T)$, which is a quadratic function describing the total level of confidence of the system. The discrete version of the Kalman Filer operates through two steps:

1. A *predictor* step, where \mathbf{u} and P are advanced in time using information derived via model only:

$$\hat{\mathbf{u}}_{k|k-1} = \Psi_k \hat{\mathbf{u}}_{k-1} + B_k \mathbf{c}_k \quad (9)$$

$$P_{k|k-1} = \Psi_k P_{k-1|k-1} \Psi_k^T + Q_k \quad (10)$$

2. An *update* step, when the observation is integrated through a weighted

measure provided by the *Kalman Gain* operator K_k :

$$S_k = H_k P_{k|k-1} H_k^T + R_k, \quad K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (11)$$

$$\hat{\mathbf{u}}_{k|k} = \hat{\mathbf{u}}_{k|k-1} + K_f (z_k - H_k \hat{\mathbf{u}}_{k|k-1}) \quad (12)$$

$$P_{k|k} = (I - K_k H_k) P_{k-1|k-1} \quad (13)$$

The two-step structure of the discrete Kalman filter is reminiscent of the predictor / corrector strategy presented for the PISO algorithm in section 2.1. This feature will be exploited to derive an estimator in section 3.

3. Integrating CFD solver & Kalman filter: a DA estimator

The present estimator combines the CFD PISO scheme with the discrete Kalman Filter in order to produce a divergence free *augmented* prediction. This is obtained considering the momentum equation 3 as the *model* for the Kalman Filter, while equation 6 is used to regularize a target velocity field. The model articulates in three macro-steps, which perform different operations whether observation is provided at the current time step or not:

1. **Predictor step.** At the beginning of the time step k , the momentum equation 3 and the error covariance matrix P time advancement in equation 10 are performed, whether or not observation is provided. This procedure is equivalent to the *update* step of the Kalman Filter. The matrix Ψ_k can be exactly obtained by manipulation of the operator Φ , which includes a matrix inversion. This problematic aspect will be discussed in detail when the reduced-order strategies will be presented. If observation is available, the Kalman gain K is calculated using equation 11.
2. **Corrector step** If observation is not available, the corrector step exactly replicates the PISO algorithm:

$$\nabla \cdot \left(\frac{\nabla p}{a_P} \right)_f = \sum_f S \times \left(\frac{\Phi(\mathbf{u})}{a_P} \right)_f \quad (14)$$

$$\mathbf{u} = \frac{\Phi(\mathbf{u})}{a_P} - \frac{\nabla p}{a_P} \quad (15)$$

These two equations are solved iteratively, until the model solution converges towards a zero-divergence condition. If observation is available, the Poisson equation is instead used to impose the zero-divergence condition for the *augmented* prediction:

$$\hat{\mathbf{u}} = \mathbf{u} + K (\mathbf{z} - H \mathbf{u}) = \frac{\Phi(\mathbf{u})}{a_P} - \frac{\nabla p}{a_P} + F \quad (16)$$

$$\nabla \cdot \left(\frac{\nabla p}{a_P} \right)_f = \sum_f S \times \left(\frac{\Phi(\mathbf{u})}{a_P} + F \right)_f \quad (17)$$

$$\mathbf{u} = \frac{\Phi(\mathbf{u})}{a_P} - \frac{\nabla p}{a_P} \quad (18)$$

Similarly to the classical version of the corrector, the loop of equations 16, 17 and 18 is performed until convergence. The resulting algorithm assures that the augmented velocity $\hat{\mathbf{u}}$ always respect the zero-divergence constraint, while this condition is observed for the velocity field \mathbf{u} derived by model only if $\nabla \cdot F = 0$. However, re-normalization of the observation in order to force it complying with the zero-divergence condition [18] is not needed here.

3. **Regularization** step. This last procedure is triggered only when observation is available. First, the error covariance matrix P is updated via equation 13. In addition, the augmented state $\hat{\mathbf{u}}$ is used to improve the performance of the numerical model. Within the framework of this research activity, the most simple solution of imposing $\mathbf{u} = \hat{\mathbf{u}}$ is chosen. In this way, the model will use the augmented velocity as a source term \mathbf{u}_0 at the beginning of the next time step.

The analysis of the proposed algorithm reveals that the classical PISO algorithm is modified only if observation is available, otherwise a classical numerical resolution is obtained. However, the resolution of the full algorithm presented is problematic when considering large dynamic systems such as those needed for the analysis of turbulent flow configurations. The principal reasons are:

1. The derivation of the matrix Ψ from the operator Φ is expensive, as it demands a matrix inversion each time observation is provided. Consider-

ing a discretization in N mesh elements, the size of the matrix should be $3N \times 3N$ because of the interactions of the three velocity components. Thanks to the linearization of the PISO algorithm, this problem is here reduced to three $N \times N$ matrices. This operation can be performed only once if the matrix Φ is constant in time, such as for laminar flows with an explicit treatment of the inertia. However, this is clearly not feasible for numerical simulation of turbulent flows, which is the target of the present research work.

2. Similar critical issues must be solved for the resolution of the error covariance matrix P . Taking benefit from the linearization of the PISO algorithm, the procedure consists of eight $N \times N$ matrix product and a matrix inversion for each velocity component. This operation demands computational resources and RAM availability which are orders of magnitude larger than the costs associated with the numerical simulation.
3. The structure of the matrix Q is difficult to predict. Suzuki [22] proposed to estimate its value using reference experimental data. While his pioneering development is remarkable, this approach excludes the bias associated with the experiments and, more in general, with the observation, which is not always true. In addition, the quantification of the level of confidence of the numerical simulation should be ideally derived accounting for model results only. Studies in the optimal control field [26] indicate that the most suitable structure for the matrix Q is diagonal. This implies that the uncertainty generated by the model on a discrete element is local and it is not initially tied to the other elements. Similar considerations are usually provided for the observation covariance matrix R . This hypothesis is reasonable if the considered model is a numerical solver, and it will be used in the present framework.

In the following, approximated solutions for the three problematic aspects introduced are proposed. The goal is to produce an efficient reduced method, which provides an optimized flow prediction with moderate increase of compu-

tational resources when compared to classical numerical simulation.

3.1. Sequential observer

The first reduced model presented is based on very strong hypothesis based on the behavior of the model matrix Ψ . During the update step, the Kalman Filter propagates the model uncertainty introduced by the matrix Q using the correlation between state variables provided by Ψ . This information is stored in the error covariance matrix P . The general formulation of the Kalman Filter is efficient for every structure of the matrix Ψ . However, applications in fluid mechanics exhibit diagonal-dominant state matrices. This is particularly true for the analysis of turbulent configurations, where a very small time step Δt has to be imposed to capture the dynamics of the flow. In addition, small time steps are required for application of the PISO algorithm, as previously discussed. Thus, while the matrix P is technically a full matrix, it shows a large number of elements ≈ 0 corresponding to the zeros of the matrix $\Psi + \Psi^T$ in the applications currently investigated.

So, the matrix Ψ is here approximated as:

$$\Psi_{OB} = \frac{1}{a_P} \Phi_0(\mathbf{u}_0) = \frac{1}{a_P \Delta t} I \quad (19)$$

The second equality is valid for the first order Euler time discretization, and it is reported for sake of clarity. This approximation implies that, if P is initially set to be diagonal and Q and R are diagonal during the simulation, the matrix P will be diagonal as well. It also implies that the uncertainty produced by the model at each time step is not propagated to other elements of the mesh. This family of estimators is usually referred to as *observers* [27, 28]. The resulting Kalman gain K_{OB} (which is as well a diagonal matrix) is sub-optimal, as the information is not shared to neighbors cells by equation 16. However, this drawback is mitigated by the resolution of the Poisson equation 17, which instantaneously propagates the information through all the physical domain. In addition, as all the matrices describing the Kalman procedure are diagonal, all the matrix products in equations 10-13 can be replaced by local algebraic

relations for every mesh element. This implies that the cost associated with the Kalman filter becomes negligible with respect to the resources required by the solver for the resolution of the momentum equation and the Poisson equation. However, integrating observation in the PISO loop usually demands a higher number of iterations to reach the prescribed level of converge of the solution. An augmentation of the 10% - 15% of the computational cost with respect to classical CFD has been observed in the practical application to the test cases investigated.

3.2. Sequential estimator based on filtering of the matrix P

A second reduced order model based on the Kalman Filter is here proposed, which similarly exploits characteristics of the dynamic system. It will be referred to as *filtering* estimator. In this case, a filtered error covariance matrix \tilde{P} is used for the computations. The only elements stored are the non-zero elements of the matrix $\Psi + \Psi^T$. In the case of three-dimensional simulations using hexahedral elements, \tilde{P} is epta-diagonal, as each mesh cell communicates with a maximum of six neighbors. Within this framework, we propose approximated formulae for matrix /matrix product and matrix inversion, which dramatically reduce the computational demands for these expensive operations. For sake of clarity the operation is now detailed for a penta-diagonal matrix, which may correspond to a two-dimensional simulation. Extension to epta-diagonal matrix is direct. Let us consider two penta-diagonal matrices A and B , which are both diagonal-dominant. The matrix \tilde{C} , which is the resulting penta-diagonal matrix obtained filtering the result of the product $C = A \times B^T$, can be approximated such that:

$$\begin{aligned}\tilde{c}_1^i &= a_1^i b_3^i + a_3^i b_1^i \\ \tilde{c}_2^i &= a_2^i b_3^i + a_3^i b_2^i \\ \tilde{c}_3^i &= \sum_{j=1}^5 a_j^i b_j^i \\ \tilde{c}_4^i &= a_4^i b_3^i + a_3^i b_4^i \\ \tilde{c}_5^i &= a_5^i b_3^i + a_3^i b_5^i\end{aligned}$$

Where the subscripts 1, 2, \dots , 5 represent the five diagonals stored for the filtered matrix. a_j^i represents the element of the matrix A located in line i and in the j^{th} diagonal. In this product, each line is independent of the others, making it appropriate for parallelization and reducing the total number of operations from $N \times N$ to $5 \times N$. If we now consider a matrix $C = I$, it is also possible to define an approximation of the right inverse of a matrix A via the resolution of a simplified system for each line ($X^T = A^{-1}$):

$$\begin{aligned}
a_1^i x_3^i + a_3^i x_1^i &= 0 \\
a_2^i x_3^i + a_3^i x_2^i &= 0 \\
\sum_{j=1}^5 a_j^i x_j^i &= 1 \\
a_4^i x_3^i + a_3^i x_4^i &= 0 \\
a_5^i x_3^i + a_3^i x_5^i &= 0
\end{aligned}$$

This approximation allows for very fast calculations of large matrices with an approximation error that decreases the more the matrix is diagonal-dominant. In order to practically illustrate the procedure, two penta-diagonal random matrix (invertible and strictly diagonal dominant) A and B are generated. Their exact product $A \times B = C$ is as well computed:

$$\begin{aligned}
A &= \begin{pmatrix} 8.02e-01 & 9.85e-02 & 0.00e+00 & 1.17e-01 & 0.00e+00 \\ 1.02e-01 & 7.38e-01 & 8.91e-02 & 0.00e+00 & 1.01e-01 \\ 0.00e+00 & 9.80e-02 & 9.43e-01 & 1.00e-01 & 0.00e+00 \\ 1.04e-01 & 0.00e+00 & 9.55e-02 & 1.02e+00 & 9.61e-02 \\ 0.00e+00 & 8.54e-02 & 0.00e+00 & 1.11e-01 & 9.10e-01 \end{pmatrix} \\
B &= \begin{pmatrix} 1.04e+00 & 1.04e-01 & 0.00e+00 & 1.03e-01 & 0.00e+00 \\ 9.90e-02 & 1.14e+00 & 1.04e-01 & 0.00e+00 & 9.93e-02 \\ 0.00e+00 & 9.45e-02 & 8.57e-01 & 1.19e-01 & 0.00e+00 \\ 1.06e-01 & 0.00e+00 & 9.11e-02 & 1.01e+00 & 9.54e-02 \\ 0.00e+00 & 8.96e-02 & 0.00e+00 & 1.10e-01 & 9.56e-01 \end{pmatrix} \\
A \times B = C &= \begin{pmatrix} 8.57e-01 & 1.96e-01 & 2.09e-02 & 2.01e-01 & 2.10e-02 \\ 1.79e-01 & 8.71e-01 & 1.53e-01 & 3.21e-02 & 1.69e-01 \\ 2.03e-02 & 2.01e-01 & 8.28e-01 & 2.14e-01 & 1.93e-02 \\ 2.16e-01 & 2.84e-02 & 1.75e-01 & 1.06e+00 & 1.89e-01 \\ 2.02e-02 & 1.79e-01 & 1.90e-02 & 2.12e-01 & 8.89e-01 \end{pmatrix}
\end{aligned}$$

The filtered product of A and B is:

$$\tilde{C} = \begin{pmatrix} 8.34e-01 & 1.03e-01 & 0.00e+00 & 1.22e-01 & 0.00e+00 \\ 1.86e-01 & 8.62e-01 & 1.78e-01 & 0.00e+00 & 1.15e-01 \\ 0.00e+00 & 1.70e-01 & 8.28e-01 & 1.84e-01 & 0.00e+00 \\ 1.05e-01 & 0.00e+00 & 2.09e-01 & 1.05e+00 & 2.19e-01 \\ 0.00e+00 & 8.16e-02 & 0.00e+00 & 1.06e-01 & 8.70e-01 \end{pmatrix}$$

And the filtered \tilde{B} obtained as $\tilde{B} = A^{-1} \tilde{\times} \tilde{C}$:

$$\tilde{B} = \begin{pmatrix} 1.04e+00 & 8.38e-02 & 0.00e+00 & 1.08e-01 & 0.00e+00 \\ 1.17e-01 & 1.15e+00 & 1.25e-01 & 0.00e+00 & 1.07e-01 \\ 0.00e+00 & 6.85e-02 & 8.50e-01 & 7.61e-02 & 0.00e+00 \\ 9.86e-02 & 0.00e+00 & 1.36e-01 & 1.02e+00 & 1.17e-01 \\ 0.00e+00 & 7.28e-02 & 0.00e+00 & 8.96e-02 & 9.52e-01 \end{pmatrix}$$

Size N	$\frac{\ C-\tilde{C}\ _F}{\ C\ _F}$	$\frac{\ B-\tilde{B}\ _F}{\ B\ _F}$
5	1.03e-01	1.93e-02
10	9.23e-02	2.27e-02
20	7.97e-02	2.33e-02
40	7.99e-02	2.91e-02
80	7.39e-02	2.82e-02
160	6.68e-02	2.69e-02
320	6.71e-02	2.89e-02
640	6.77e-02	2.84e-02
1280	6.62e-02	2.90e-02

$A_{ii}/A_{ij}, i \neq j$	$\frac{\ C-\tilde{C}\ _F}{\ C\ _F}$	$\frac{\ B-\tilde{B}\ _F}{\ B\ _F}$
5	1.73e-01	6.97e-02
10	6.68e-02	2.77e-02
100	4.92e-03	2.81e-03
1000	4.97e-04	2.78e-04

Table 1: Frobenius norm of the relative error committed with the approximated matrix product. On the left, the sensitivity of the error to the size of the matrix is investigated. In this case the parameter $A_{ii}/A_{ij}, i \neq j$ measuring the dominance of the diagonal terms is set to 10. On the right, the sensitivity of the approximation error to the parameter $A_{ii}/A_{ij}, i \neq j$ is studied, fixing a matrix size of 640×640 .

Observation of the previous results proves that the filtered matrices \tilde{B} and \tilde{C} are reasonable approximations of the exact matrices B and C . The analysis has been extended to larger matrices and the relative error has been quantified via Frobenius norm $\|\cdot\|_F$. The results reported in table 1 indicate that the approximation error committed is not sensitive to the size of the matrix, but as expected it decreases with increasing dominance of the diagonal terms. In the case of numerical simulation of turbulent flows, the parameter $A_{ii}/A_{ij}, i \neq j$ measuring the dominance of the diagonals terms is usually of the order of 5 – 10.

In summary, the *filtering* estimator resolves the full scheme presented in Section 3 via the use of the approximated matrix products / inversions here discussed. While the resulting Kalman gain is sub-optimal, the level of precision is supposedly higher than the diagonal approximation of the *estimator* observer presented in Section 3.1, because information transmitted between neighbors is here conserved. The Authors would like to stress two important points:

1. Even if differences of the order of 10% can be measured by the comparison of \tilde{K} and K , this does not absolutely imply that a similar level of mag-

nitude of difference is observed in the flow field predictions. The Kalman gain is actually a gain, and it just approximately regulates the amount of observation that is included in the model prediction. Additionally, K is an optimal gain with respect to a prescribed quadratic norm. While this criterion is robust and reliable, it can not be considered as an exact quantification of the quality of the resulting flow prediction. Tests on simplified test cases in Section 4 have shown that negligible differences are observed when applying a full Kalman Filter, an observer estimator or a filtering estimator.

2. The structure of the estimator presented in Section 3 is reminiscent to implementation of porous sources or discrete Immersed Boundary Method (IBM) forcing [29, 30]. The term $K(\mathbf{z} - H\mathbf{u})$ actually forces the predicted flow towards the asymptotic limit \mathbf{z} and is regulated in intensity by the Kalman gain K . In classical discrete IBM application, we would simply have $K = 1$. In addition, the information is integrated locally and then propagated by the pressure field in the *observer*. On the other hand, the observation is spread to the cell neighbors through interpolation (governed by K) in the *filtering* estimator, before propagation via the Poisson equation.

In the following, it will be shown that the *augmented prediction* of turbulent flows can be performed with present computational capabilities, and it promises groundbreaking advance in CFD.

4. Manufactured tests cases

The analysis of two classical test cases via estimator is here proposed. The dynamics are not represented by the Navier–Stokes equations, but simplified equation of diffusion and advection are considered. Because of this aspect, the computational resources required to analyze these cases are very small and they allow for a complete investigation using the full Kalman filter and the two reduced order model proposed. In addition, an exact solution is known, which

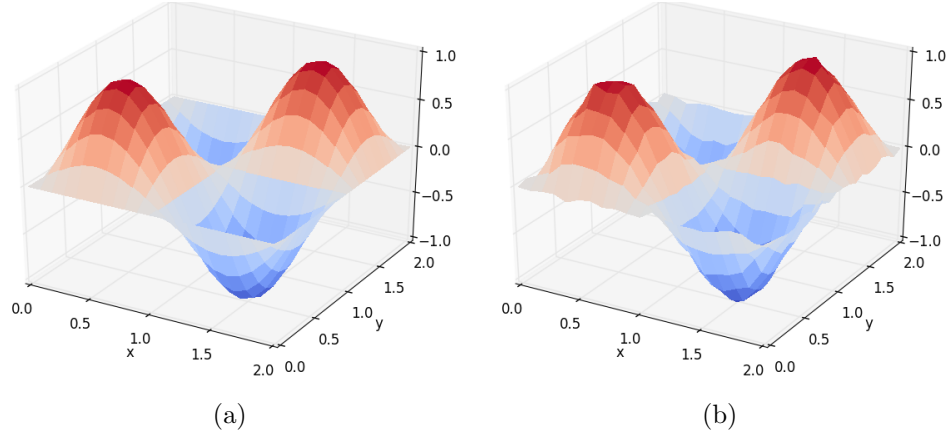


Figure 1: Diffusion of a scalar T : (a) manufactured solution and (b) *observation* for $t = 1$.

will be perturbed via controlled noise to produce suitable observation and to provide an exact estimation of the covariance matrix R .

The first test case investigated is a pure diffusion problem of a scalar T in the two-dimensional domain $[0; 2] \times [0; 2]$, which is illustrated in figure 1:

$$\partial_t T - \Delta T = g(t) \quad (20)$$

ΔT is here the laplacian of T and Homogeneous Dirichlet boundary conditions are imposed. A suitable source term $g(t)$ serves a manufactured solution which is also used to set a non-zero initial value:

$$T_{ref}(x, y, t) = \sin(\pi x) \sin(\pi y) \cos(2\pi t) \quad (21)$$

The numerical discretization of this case is performed via a finite volume scheme of second order precision in space and explicit Euler (first order) in time.

The second test case is a pure advection problem of a scalar ϕ in the two-dimensional domain $[-10; 10] \times [-10; 10]$, which is shown in figure 2:

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0 \quad (22)$$

with homogeneous Dirichlet boundary conditions. A non-zero initial value for ϕ is imposed:

$$\phi_{ref}(x, y, t) = \max \left[0, 1 - \left(x - 5 \cos\left(\frac{\pi t}{10}\right) \right)^2 - \left(y - 5 \sin\left(\frac{\pi t}{10}\right) \right)^2 \right] \quad (23)$$

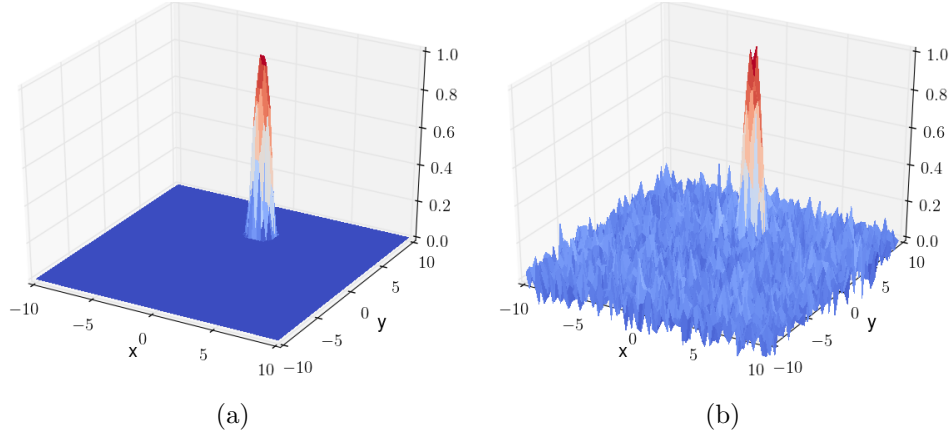


Figure 2: Transport of a scalar φ : (a) manufactured solution and (b) *observation* for $t = 5$.

The advection velocity is defined as $\mathbf{u} = \frac{\pi}{10}(x\vec{e}_y - y\vec{e}_x) = r\frac{\pi}{10}\vec{e}_\theta$, where $(\vec{e}_r, \vec{e}_\theta)$ define a cylindrical coordinate system. The numerical discretization used for this case is a first order upwind scheme for the spatial derivatives and explicit Euler in time. This discretization scheme is known to be diffusive. This feature will be exploited in order to magnify the quality improvement prediction obtained using the Kalman Filter.

In both cases, the *observation* is obtained via perturbation of the manufactured solutions imposing a Gaussian noise of variance σ_r . As a consequence, the observation covariance matrix is $R = \sigma_r I$. The model is starting using the manufactured solution as initial condition, so that the error covariance matrix is $P = 0$. However, information about the structure of the model covariance matrix Q is not determined yet. For this reason, it is initially hypothesized that $Q = \sigma_q I$. A number of test has been performed, choosing different values of the mesh resolution and of the parameters σ_r and σ_q . Results are reported in figure 3 for the advection test case and in table 2 for both test cases. In the table, L2 norms are calculated comparing the exact solution with the observation, the model and the prediction by Kalman filter, respectively. The error for the observation is clearly driven by the parameter σ_r , while the mesh resolution affect the accuracy of the model prediction. The configuration obtained via

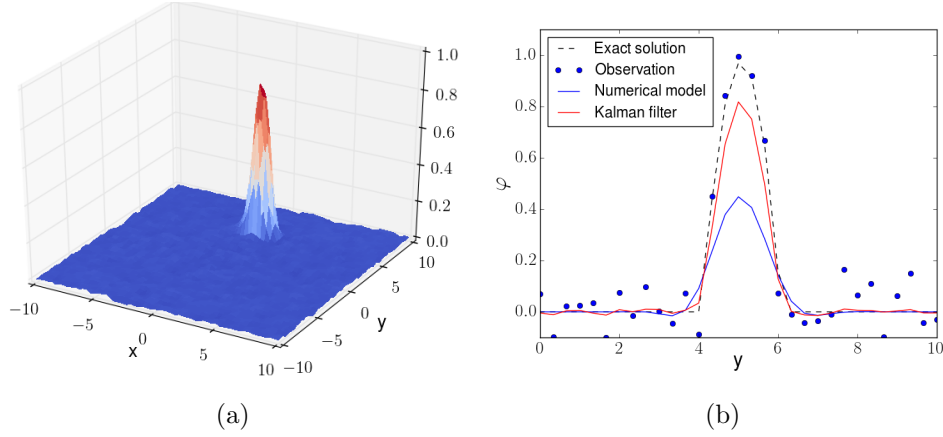


Figure 3: Pure advection of a scalar φ . (a) Numerical solution provided by application of Kalman filter for $t = 1$. (b) Solution for $x = 0$ and $t = 5$.

Kalman Filter, which is sensitive to all of the three parameters, is potentially the most accurate. This is particularly true when the level of confidence in the observation and in the model is similar. In addition, the analysis of figure 3 indicates that Kalman Filter application can reduce drawback and biases associated with the model, such as the numerical diffusion of the upwind schemes using for numerical discretization.

While the definition of the covariance matrix R is straightforward, the determination of the matrix Q is problematic. The covariance coefficient σ_q is now optimized targeting the minimization of the difference between the prediction via Kalman Filter and the exact solution. This procedure provides insights about the structure of the model covariance matrix Q , which is still largely unexplored in CFD applications. The optimization algorithm is the gradient descent and three different test have been performed: i) confidence in the observation and the model roughly the same, ii) more confidence in the observation and ii) more confidence in the model. Results are shown in figure 4. The most important conclusion that can be drawn is that there is a linear correlation between the optimized value of σ_q and the variance coefficient σ_r . This observation origin from the behavior of the Kalman Filter, which is optimized when

Mesh size	σ_r	σ_q	Error		
			Observation	Model	Kalman Filter
1.05e-01	5.91e-05	9.77e-07	2.17e-02	1.09e-02	4.56e-03
1.05e-01	1.48e-05	9.77e-07	1.09e-02	1.09e-02	2.53e-03
1.05e-01	3.69e-06	9.77e-07	5.43e-03	1.09e-02	2.08e-03
6.90e-02	7.00e-06	5.41e-07	7.48e-03	7.47e-03	1.63e-03
2.00e-01	1.69e-03	6.36e-05	1.49e+00	7.46e-01	3.48e-01
2.00e-01	4.22e-04	6.36e-05	7.45e-01	7.46e-01	2.57e-01
2.00e-01	1.05e-04	6.36e-05	3.73e-01	7.46e-01	1.99e-01
1.50e-01	2.20e-04	2.15e-05	7.00e-01	7.02e-01	2.21e-01

Table 2: L2-norm of the error obtained for the observation, the model and the prediction via Kalman Filter when compared to the exact solution. Parametric analysis of the error with respect to the observation covariance σ_r , the model covariance σ_q and the mesh size is shown. (top) The diffusion of a scalar T and (bottom) the advection of a scalar ϕ are reported, respectively. Due to the randomness of the Gaussian noise applied to the observation, the error has been averaged over a significantly large number of independent runs.

the level of confidence in the observation and model are roughly the same. The sensitivity of this algorithm has been extensively tested changing the mesh size and the time step Δt , as shown in figure 5 for the pure diffusion test case. One notable property is that the optimized value of σ_q is directly linked with the truncature error of the numerical scheme multiplied by Δt . In the case of the diffusion equation, this is equal to:

$$\Delta t E_T = \Delta t^2 f_t \left(\frac{\partial^2 T}{\partial t^2} \Big|_P \right) + \Delta t \Delta x^2 f_s \left(\frac{\partial^4 T}{\partial x^4} \Big|_P, \frac{\partial^4 T}{\partial y^4} \Big|_P \right) \quad (24)$$

where f_t , f_s are linear functions of higher order derivatives of the scalar T , calculated in the mesh element P . Three cases are analyzed in figure 5. For the first one, a coarse mesh of $N \times N = 11 \times 11$ elements has been used. In this case, the optimized value of σ_q scales as $\Delta t^{3/2} \cdot \Delta x^2$, which is a hybrid behavior of the two sources of error in equation 24. Arguably, boundary effects govern the numerical prediction, considering that 33% of the mesh is composed by boundary elements. However, the early emergence of two distinct regimes

can be observed for the case $N \times N = 31 \times 31$ (12.4% of boundary elements). For high Δt values, the optimized value of σ_q scales as Δt^2 i.e. error due to the discretization of the time derivative. As the time step is refined, the overall contribution of the time discretization becomes less and less dominant. It appears that the limit $\sigma_q \propto \Delta t \cdot \Delta x^2$ is approached increasing the space resolution. Thus, a transition towards a state where Q is linked with the space discretization error is observed.

In conclusion, the results presented in figure 5 actually indicate that the level of confidence in the numerical model is related with the discretization error. This trend of correlation is logical, because the lack of precision in the numerical results, which represents the uncertainty of the model, is locally produced by the discretization error. These findings are important because they indicate that the model uncertainty can be at least estimated using model information only. The functions f_t and f_s can be locally calculated for each mesh element, when a suitable solution is provided. In section 6 the determination of the structure of Q for the analysis will be simplified, envisioning direct application without any preliminary information about the flow field.

At last, the performances of the full Kalman filter, the filtering estimator and the observer have been assessed via comparison for the test cases presented in this section. The analysis has encompassed a large spectrum of configurations, including different mesh resolutions and values of σ_r and σ_q . For the latter, both optimized and random values have been considered. The main results, which are reported in table 3, indicate that three approaches are almost equivalent. In addition, optimized values of σ_q for the three approaches are very similar as well, reinforcing the idea that the reduced-order methods can provide similar qualitative performance of the full Kalman filter.

5. Application to laminar test cases: flow around a circular cylinder

Applications of the reduced order estimator to the analysis of laminar flow configurations are investigated in the presents Section. The flow around a cir-

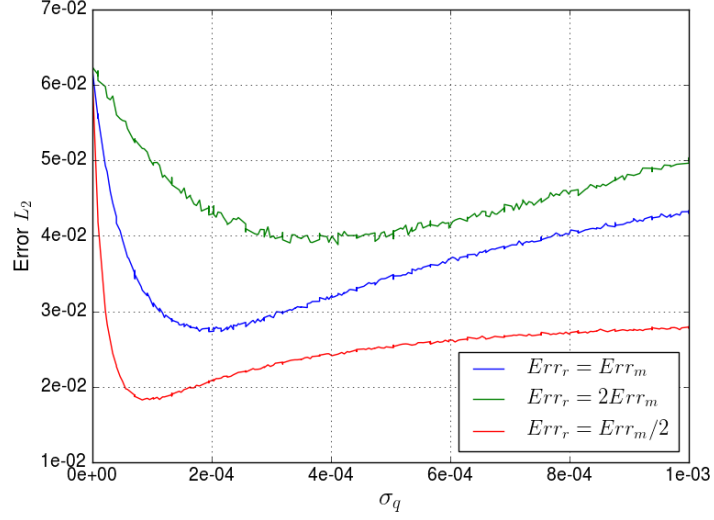


Figure 4: Sensitivity of L2 norm of the error of the Kalman Filter prediction as a function of σ_q . $Err_r = Err_m$ correspond to a choice of σ_r such as the L2 norm the error of the observation is roughly the same as the L2 norm of the pure simulation. The two other cases correspond respectively to a σ_r twice as bigger or smaller.

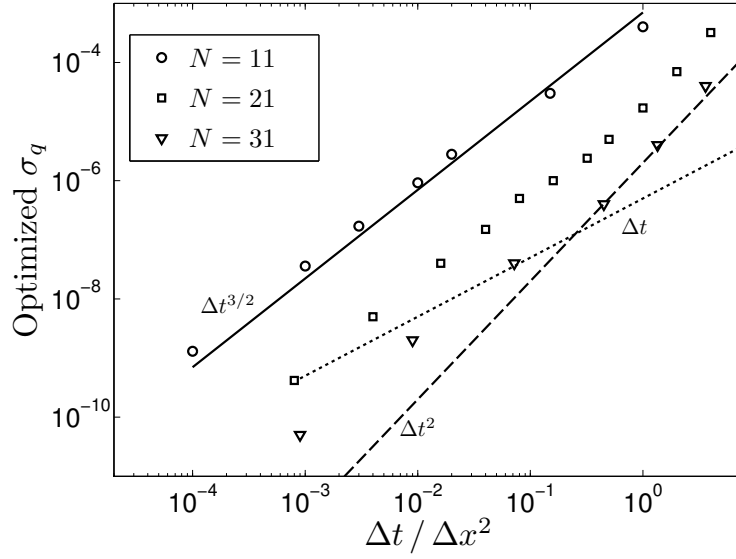


Figure 5: Sensitivity of the optimized value of the model variance σ_q to Δt , Δx . The test case of 2D diffusion is investigated.

Mesh size	σ_r	σ_q	Error		
			full Kalman	filtering estimator	observer
1.05e-01	4.86e-04	2.04e-04	2.51e-02	2.74e-02	2.76e-02
1.05e-01	4.86e-04	1.57e-04	2.38e-02	2.67e-02	2.86e-02
1.05e-01	4.86e-04	1.18e-04	2.28e-02	2.88e-02	3.04e-02
6.90e-02	4.69e-04	2.10e-04	2.37e-02	2.64e-02	2.66e-02
3.00e-01	9.49e-04	1.78e-04	3.14e-01	3.10e-01	3.21e-01
2.00e-01	4.22e-04	5.12e-05	2.59e-01	2.97e-01	3.10e-01
2.00e-01	4.22e-04	5.63e-05	2.93e-01	2.56e-01	3.01e-01
2.00e-01	4.22e-04	5.74e-05	2.91e-01	2.88e-01	2.68e-01

Table 3: L2-norms of the error observed for the diffusion case comparing full Kalman Filter, filtering estimator and observer to the exact solution. Parametric analysis of the error with respect to the observation covariance σ_r , the model covariance σ_q and the mesh size is shown. Values in bold text correspond to cases where σ_q is optimal. Due to the randomness of the Gaussian noise applied to the observation, the error has been averaged over a significantly large number of independent runs.

cular cylinder for $Re = 100$ is considered. This simple geometric case allows for the observation of an unstationary Von-Karman vortex street. In addition, it is very well documented in the literature (see [31, 32, 33, 34, 35]). The analysis is here focused on the application of the *observer* introduced in Section 3.1.

5.1. Two-dimensional test case

This flow configuration exhibits a two-dimensional unstationary behavior at $Re = 100$, so that the numerical domain investigated is initially 2D. The physical quantities of reference are normalized, so that the asymptotic streamwise velocity in the direction x is $\mathbf{u}_\infty = 1$ and the diameter of the cylinder is $D = 1$. The resulting normalized kinematic viscosity is derived as $\nu = \mathbf{u}_\infty D / Re = 10^{-2}$.

The physical domain investigated is circular and defined by a diameter of $D_c = 30$. The inflow boundary conditions imposed is a Dirichlet condition for velocity $\mathbf{u} = \mathbf{u}_\infty$. A mass-conserving advective condition is imposed at the outlet. A cylindrical coordinate system is chosen to perform the mesh discretization. The physical domain is discretized in 140 cells in the azimuthal

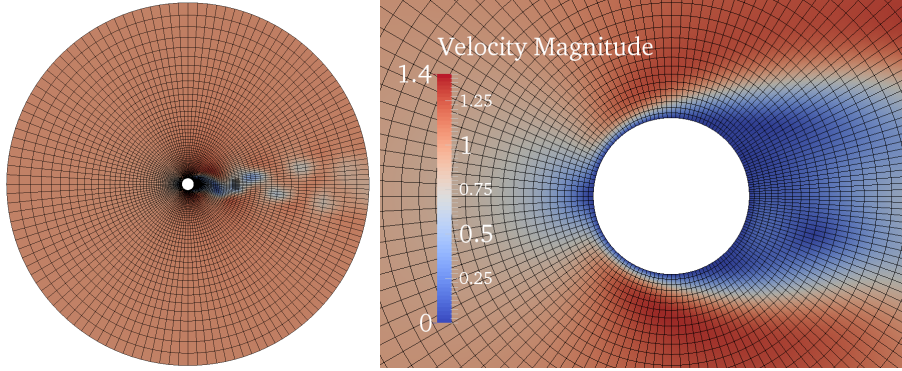


Figure 6: Computational domain investigated for the flow around a circular cylinder, $Re = 100$.

direction and 50 mesh elements in the radial direction, for a total of 7×10^3 elements (see figure 6). A geometric distribution of the mesh elements has been chosen in the radial direction, in order to obtain a finer resolution at the wall. A constant time step has been chosen for time advancement with $\Delta t = 0.03t_A = 0.03 D/\mathbf{u}_\infty$. This corresponds to a maximum Courant number of $C_O = 1.22$.

The aerodynamic coefficients investigated are defined as:

- The lift coefficient $C_L = \frac{2F_y}{\rho u_\infty H_D}$
- The drag coefficient $C_D = \frac{2F_x}{\rho u_\infty H_D}$

where F_y and F_x are the aerodynamic forces in the normal direction (y) and in the streamwise direction (x). They include both pressure and viscous contributions. The surface H_D represents the area of the cylinder normal to the streamwise direction. In addition $\rho = 1$ in the normalized system. A first test numerical simulation is performed, in order to get rid of the initial transient. A fully developed configuration is then used as initial condition for the following simulations.

The strategy adopted for Data Assimilation completely relies on CFD for this case, using the same solver to produce observation and as a model. The main goals of this analysis are to test the capabilities of the model to i) pro-

duce a smooth augmented solution and ii) influence the physical emergence of unsteady phenomena. Traditionally, sequential methods operate efficiently when observation is provided upstream. On the contrary, they under-perform when information is provided downstream, with respect to variational approaches. Thus, the observation window for this analysis consists of 24 mesh elements only, which are located in the wake area at approximately one full shedding cycle downstream (see figure 7). Observation is performed via sampling of u_x and u_y produced by a first numerical simulation. Data is collected every five time steps. The error covariance matrix P is initially set to zero, while the covariance matrices are chosen as $R = \sigma_r I$, $Q = \sigma_q I$. In this case, σ_r and σ_q are not calculated, but we impose $\sigma_r = \sigma_q$ i.e. the level of confidence in the observation is assumed to be the same as the simulation. Considering that the Data Assimilation is performed every five time steps, the diagonal values of the Kalman Gain will be slightly larger than 0.5. When the Data Assimilation is performed, the CFD model starts using the same initial configuration and solver as for the previous sampling, but the observation is introduced approximately in phase opposition starting from time $t = 10$, as shown in figure 8 for the aerodynamic coefficients. After a suitable transient, the numerical model completely adjusts the phase of its shedding cycle with the observation by global flow modification. An instantaneous velocity profile during this transition is shown in figure 9. The estimator operates like a forcing in the flow field and the Poisson equation diffuses this result in all the physical domain. This last result indicates that the sequential estimator is able to efficiently govern upstream flow dynamics when data is provided downstream in the wake. In addition, data has been voluntarily included in a small section of the physical domain, in order to reproduce most of the occurrences where observation can be only locally provided. These properties are essential for integration of experimental samples in numerical simulation.

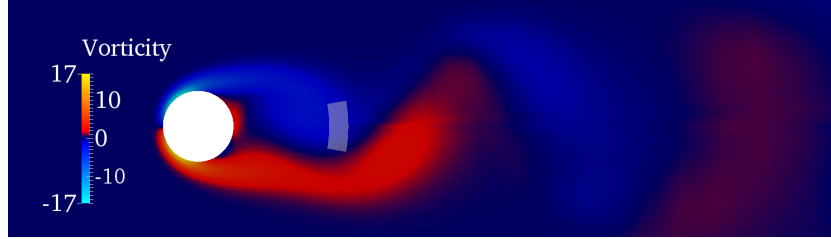


Figure 7: Isocontours of the vorticity magnitude. The shaded area corresponds to the observation window chosen for integration of data in the estimator.

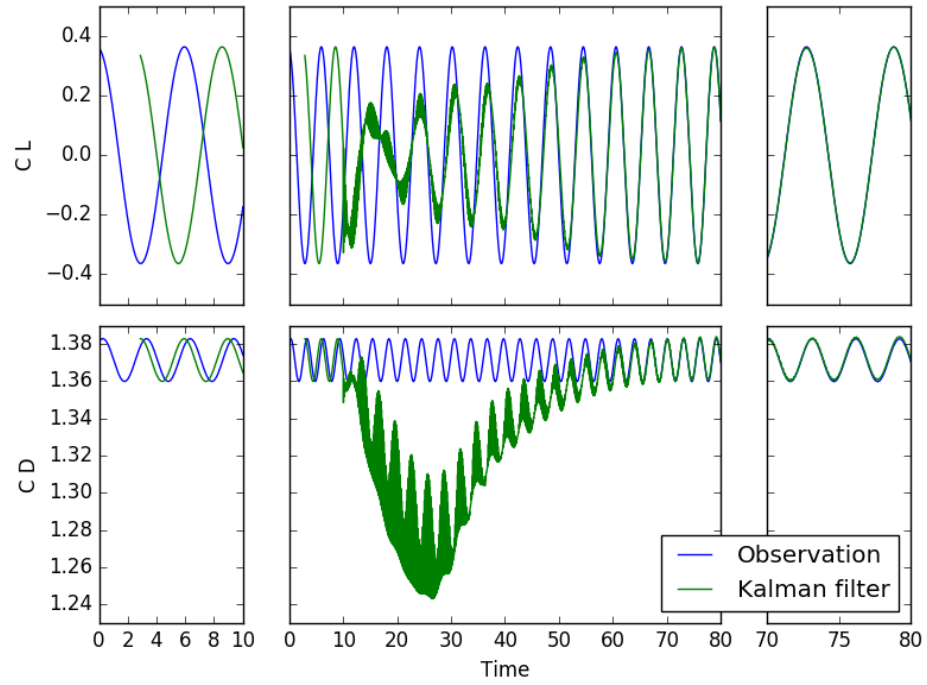


Figure 8: Time evolution of the coefficients (top) C_L (bottom) C_D during Data Assimilation for the circular cylinder in 2D.

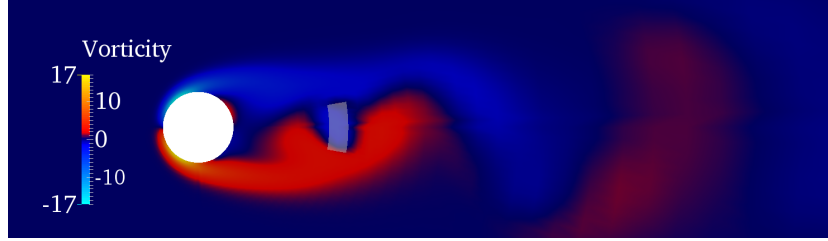


Figure 9: Instantaneous vorticity isocontours during the first phases of Data Assimilation via estimator. The model acts as a forcing while synchronizing the numerical model to the observation.

5.1.1. Three-dimensional test case

The flow around a circular cylinder for $Re = 100$ is now studied over a three-dimensional domain. The set-up of the simulation is very similar to the two case, with the following exceptions. The length of the cylinder in the spanwise direction z is $l = 30$ and it is discretized in 20 mesh elements. Periodic boundary conditions are imposed on the lateral surfaces.

The definition of the matrices P , R and Q is identical to the two-dimensional case, and the development of the analysis is exactly the same. However, a two-dimensional observation region has been conserved, which is located at the center of the domain with respect to the spanwise direction. It also is wider in the streamwise direction (around 100 cells) in order to produce a faster synchronization. The Data Assimilation process shows similar features to the correspondent two-dimensional case, as shown in figure 10 for the coefficients C_D and C_L . However, the analysis of the streamlines sampled during the transient in figure 11 indicates that the information is successfully propagated in the spanwise direction, before a complete return to a two-dimensional flow is observed after the synchronization.

6. Augmented flow prediction for turbulent flow configurations

The reduced order estimator presented in section 3 is now applied to the study of turbulent flow configurations. This goal is a fundamental research

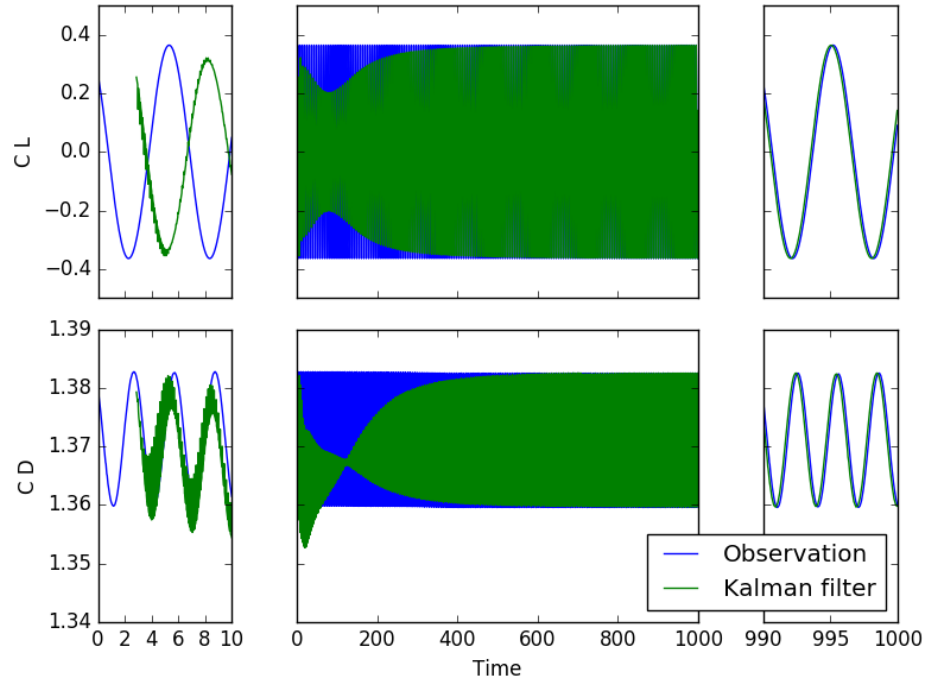


Figure 10: Time evolution of the coefficients (top) C_L (bottom) C_D during Data Assimilation for the circular cylinder in 3D.

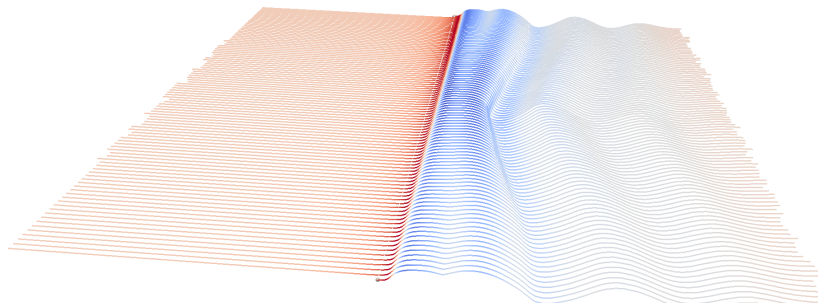


Figure 11: Streamlines just above the cylinder showing the flow becomes three-dimensional as the information is propagated by the estimator.

objective for a large spectrum of industrial applications, including transport engineering and environmental sciences. Two different applications are presented, which deal with simplified flows commonly observed in complex applications. The focus here has been on the prediction of statistical moments, which is a classical research objective in turbulence studies. In both cases, the *observer* version of the model has been used, in order to reduce the computational resources demanded. The estimator required $\approx 10\%$ resources more than the CFD model. This result indicate that real-time application for the analysis of turbulent flows are going to be possible in the next decade, in particular for application of esatechnologies resources expected in the next years. In addition, simplified formulae have been introduced for the estimation of the diagonal elements of the matrix Q . These proposals envision a simple application for an external user, without the need to provide an accurate estimation of high order gradients of the physical quantities as in equation 24. First of all, because of the need to use small time steps Δt with respect to the average mesh size $\Delta l = \sqrt[3]{\Delta V}$, the error associated with time discretization is supposed to be negligible with respect to the space discretization. For this reason, the elements of the matrix Q are locally calculated for the mesh element i as

$$Q_i = C \Delta t^{ot} \Delta l^{os} \quad (25)$$

where C is a constant chosen by the user and ot , os are related with the order of the numerical schemes used for time / space discretization, respectively. Clearly, the lower the value of C and the higher the values of ot and os , the higher the confidence level in the numerical tool used.

A second aspect that is here considered in the interaction between numerical error and turbulence modeling, which is one of the most problematic issues in particular for Large Eddy Simulation [36]. These non linear interactions are responsible for a very high level of uncertainty in the results of the numerical simulation and this challenging aspect is currently subject of investigation by numerous research teams. In the spirit of a simple approximation, equation 25

is adapted for the application to reduced-order numerical simulation based on *eddy viscosity* models [3, 4]:

$$Q_i = C \left(1 + \frac{\nu_T}{\nu} \right) \Delta t^{ot} \Delta l^{os} \quad (26)$$

where ν_T represents the turbulence / subgrid scale viscosity introduced by the model. While equation 26 does not encompass application for all the turbulence / subgridscale models reported in the literature, it can be applied to most of the models implemented in commercial / open-source codes available. The mechanics of equation 26 is simple and clear. If the volume of the mesh element i is large and the value of ν_T is high, the confidence level in the local performance of the numerical solver is severely penalized.

These techniques are now applied to the analysis of two flow configurations of industrial interest, namely the spatially evolving mixing layer and the flow around a thick plate.

6.1. Spatially evolving mixing layer, $Re_\Delta = 100$

The spatial evolution of a mixing layer [37, 38, 39, 40] is here investigated. This classical flow configuration exhibits Kelvin–Helmoltz instability structures that are observed in many industrial flows, which emergence from the interaction of the two asymptotic flows described by the velocities U_1 and U_2 , respectively. A visualization of the case investigated and isocontours of the Q criterion are reported in figure 12. If the Reynolds number is sufficiently high, the flow undergoes a turbulent transition sufficiently downstream from the inlet. The location at which the emergence of turbulent features is observed can be controlled via features of the velocity profile imposed at the inlet.

The flow configuration investigated is described by the following parameters:

- $Re_\Delta = \frac{\delta_0(U_1 - U_2)}{2\nu} = 100$, where $\delta_0 = 5 \times 10^{-4}$ is the vorticity thickness at the inlet and $\nu = 2.5 \times 10^{-5}$ is the molecular viscosity.
- $\alpha = \frac{U_1 - U_2}{U_1 + U_2} = 0.27$. This parameter, which represents the ratio of shear vs advection effects, is investigated in the literature in the range $\alpha \in$

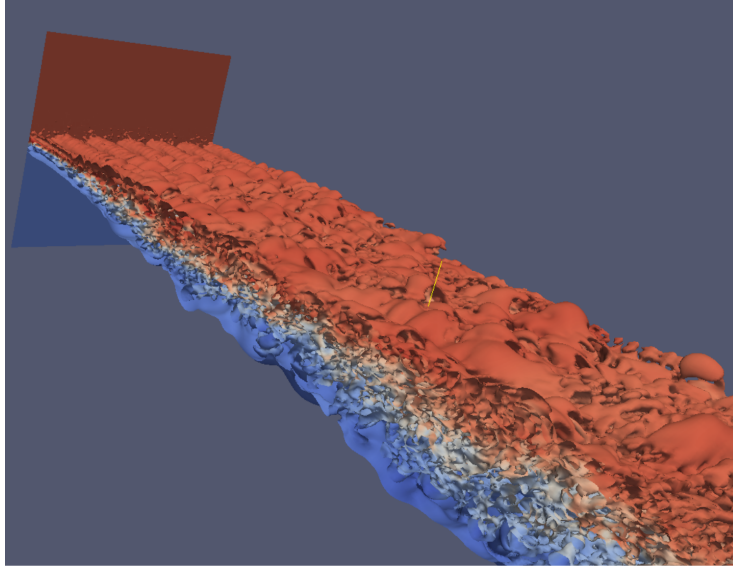


Figure 12: Spatially evolving mixing layer flow configuration, $Re_{\Delta} = 100$. Isocontours of the Q criterion colored by the velocity field, obtained by DNS. The flow exhibit a progressive transition towards turbulence moving downstream.

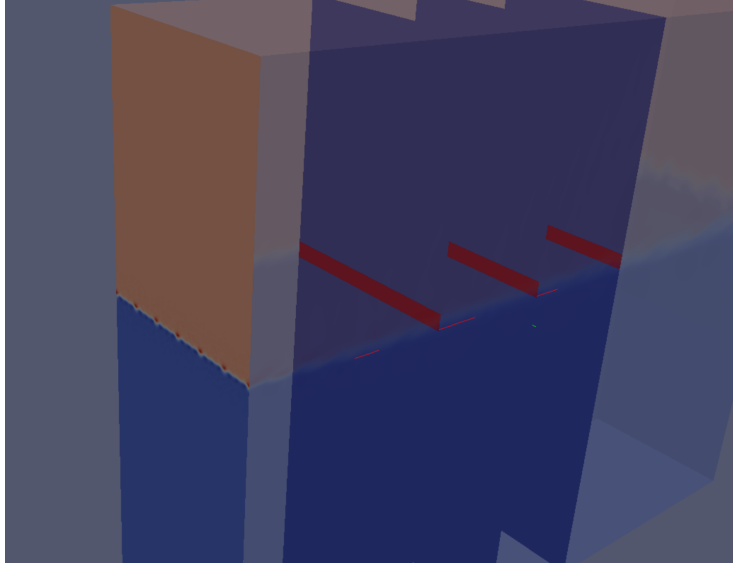


Figure 13: Spatially evolving mixing layer flow configuration. Instantaneous velocity profile imposed at the inlet, as well as the planes where observed data are provided (the exact location is highlighted in red).

[0.2, 0.7]. A value of α closer to the lower end has been chosen here in order to limit to a minimum the interaction of the wake with the boundary conditions in the normal direction.

The application of the estimator is completely numerical in this case. In fact, the tools used to derive the augmented prediction are:

- An LES solver as a *model*. Subgrid scale modeling is performed via the Smagorinsky model [41].
- DNS samples have been used as *observation*.

These tools are integrated in the *observer* estimator, which will be referred to as DA-LES. The research activity is justified by the analysis of the performance of the Smagorinsky model, which is over-dissipative at the large scales of the flow [42, 43]. In the case of the spatially evolving mixing layer, this features results in an over-prediction of the momentum thickness

$$\bar{\Theta} = \int_{y_m}^{y_M} (U_1 - \overline{u_x})(\overline{u_x} - U_2) dy \quad (27)$$

where $\overline{u_x}$ is the time averaged streamwise velocity and y_m / y_M represent the lower and upper limit of the physical domain in the normal direction, respectively. The over-prediction of $\bar{\Theta}$ for the Smagorinsky model is associated with an over-dissipation acting in the transition region of the flow, which is progressively carried downstream in the turbulent wake. The underlying idea is to assess the capability of the estimator to correct this unfavorable characteristic of the Smagorinsky model via DNS observation in the transition region.

For every numerical simulation, the set-up is the following:

- The physical domain in the streamwise, normal and spanwise directions $[x, y, z]$ is $[0, 24] \times [-9, 6] \times [-3, 3]$ in $\Lambda = 15.4\delta_0$ units. This characteristic length is the principal instability length calculated via linear stability theory.
- The numerical schemes used for discretization of spatial derivatives are

centered second-order schemes. For the time derivative, a second-order implicit backward differencing scheme has been adopted.

- Periodic conditions have been imposed on the lateral boundaries (planes $x - y$), while traction-free conditions have been imposed on the normal boundaries (planes $x - z$). An advective mass conserving condition has been imposed at the outlet. An hyperbolic tangent profile for the velocity field has been imposed at the inlet. This profile can be written as the sum of three different components $u_{INLET} = u_{th} + u_{sp} + u_{tp}$ where:

1. $u_{th} = 0.5(U_1 + U_2) + 0.5(U_1 - U_2) \tanh(2y/\delta_0)$ is an hyperbolic tangent velocity profile. This component is imposed for u_x only.
2. $u_{sp} = 0.005(U_1 + U_2) \sum_{g=1}^3 \cos(2\pi g z / \Lambda) \exp(-0.5(y/\delta_0)^2)$ is a spatial modal perturbation, which is exponentially damped moving away from the center-line. This perturbation is applied to all the velocity components.
3. $u_{tp} = 0.02(2 \text{rand}(t)[-0.5, 0.5])(u_{th} + u_{sp}) \exp(-0.5(y/\delta_0)^2)$ is a white noise time perturbation of maximum intensity equal to 2%, which is exponentially damped moving away from the center-line.

The velocity profile is represented in figure 13. Using this inlet condition, the transition to turbulence is observed at a streamwise distance of $x \approx 10\Lambda$ [40].

- The DNS mesh is structured in $384 \times 163 \times 160$ elements. The distribution in the streamwise x direction and in the spanwise z direction is uniform. In the normal direction, 64 uniform mesh elements are used to represent the range $y \in [-\Lambda, \Lambda]$. Outside of this range, a local stretching ratio of ≈ 1.1 is used as the elements become coarser reaching the boundaries in the normal direction. The resulting computational mesh is composed by 10^7 elements. At the center-line, the resolution is estimated to be $[5.86\eta, 2.93\eta, 3.5\eta]$, where η is the Kolmogorov scale calculated using the asymptotic turbulent Reynolds number. For the LES simulations the mesh

is made of $96 \times 89 \times 40$ elements. The distribution pattern is very similar to the DNS strategy, but the resolution in the streamwise and spanwise direction has been reduced by a factor of 4. In the normal direction for $y \in [-\Lambda, \Lambda]$ the coarsening ratio imposed is equal to 2.

- The time step for each numerical simulation has been set to $\Delta t = \frac{t_A}{100} = \frac{\Lambda}{50(U_1+U_2)}$, where t_A is the characteristic advection time. The maximum Courant number observed is $Co \approx 0.35$ for the DNS calculation and $Co \approx 0.08$ for the LES. Each numerical simulation has been run for 20000 times steps, for a total simulation time $T = 200t_A$.
- The initial condition for every simulation is a fully developed configuration calculated by a preliminary DNS. This velocity field has been interpolated on the LES grid.

The choice of the structure of the covariance matrices Q and R is now discussed. For the LES model, equation 26 is used to determine the value of the elements of the matrix Q :

$$Q_i = C_{ML} \left(1 + \frac{\nu_T}{\nu}\right) \sqrt[3]{\Delta V^2} \Delta t \quad (28)$$

in this case, the confidence level in LES results is clearly lower in the wake region, where the Smagorinsky model introduces the eddy viscosity effect (see figure 15 (a) for time averaged isocontours). This feature will be exploited by the estimator, in order to limit the unwanted effects of the model in the transition region. A very similar structure is chosen for the elements R_i of the covariance matrix of the observation by DNS using equation 25:

$$R_i = C_{ML} \sqrt[3]{\Delta V^2} \Delta t \quad (29)$$

The DNS observation is integrated in the LES estimators in three different planes normal to the streamwise direction x for $x = 6\Lambda, 10\Lambda, 14\Lambda$. A visual representation is reported in figure 13. For each plane, the assimilation is limited to $y \in [-1.2 \times 10^{-3}\Lambda, 1.2 \times 10^{-3}\Lambda]$, $z \in [-3\Lambda, 3\Lambda]$, which includes the eight

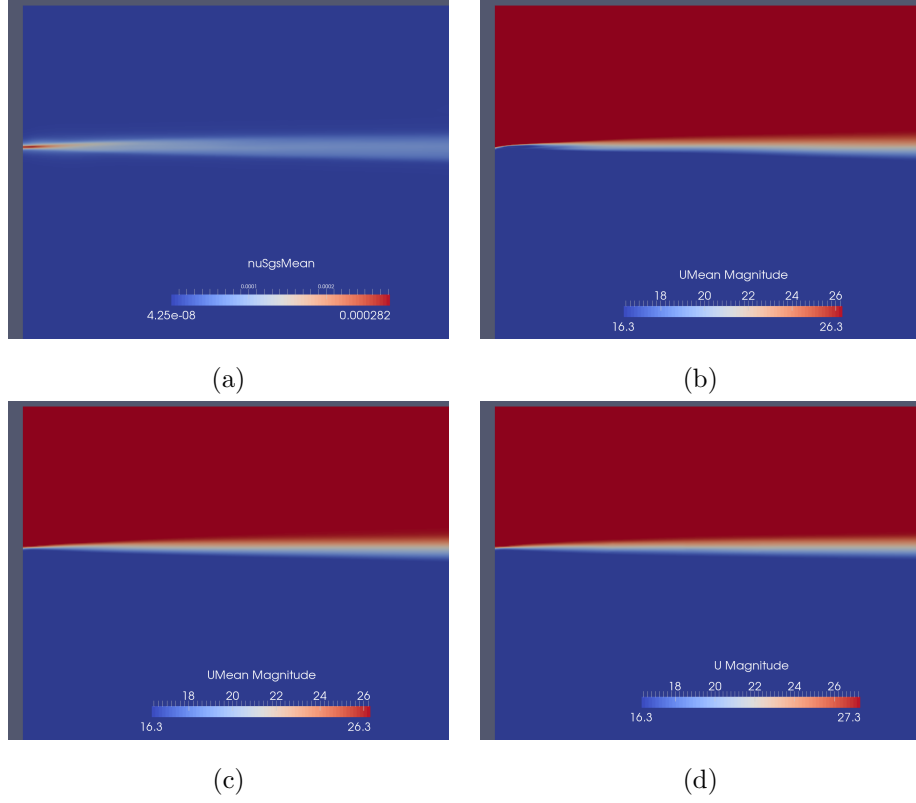


Figure 14: (a) Visualization of the averaged subgrid velocity ν_T calculated by the Smagorinsky model in the LES simulation. (b-d) Isocontours of the time average velocity $\|\bar{\mathbf{u}}\|$ on a plane normal to the spanwise direction z . The results are shown for (b) the DNS calculation, (c) the LES simulation and (d) the observer.

elements closer to the center-line in the normal direction and all the elements in the spanwise direction. Thus, the observation is provided in 960 elements of the LES mesh, which corresponds to $\approx 0.3\%$ of the total number of elements. The choice of the observation points is performed accordingly to previous considerations about the transition of the flow to a fully turbulent state. In this scenario, the DNS data is expected to improve the performance of the LES solver in the transition region, where the LES performance is supposedly sub-optimal. It is important to remind that in this case the subgrid scale model is not modified, but that the velocity field is updated comparing the LES and the DNS fields. The observation has been sampled every 10 time steps, which implies that DNS data are assimilated 2000 times during the simulation. DNS data are interpolated on the centers of the elements of the LES grid for easier application, and Data Assimilation is performed for all the three components of the instantaneous velocity field. As the initial physical field introduced in the estimator is interpolated from DNS results, the initial value of the error covariance matrix P is set to zero. Considering equations 28 and 29 and that the time step for the assimilation is 10 times larger than Δt , one can easily deduce that the diagonal elements of the optimal gain K will be very close to the unity for each velocity component. This is technically not the optimal application scenario for the estimators, which has proven to perform best when the confidence in the model and in the observation are similar. Thus, this is a good benchmark to test critical issues of the DA model when applied to the simulation of turbulent flows.

Isocontours of the time-averaged velocity magnitude are reported in figure 14 (b-d), for the three numerical simulation. The velocity profiles are very similar, but it can be observed that the transition between the two asymptotic regimes seems to be somehow faster for the LES results in figure 14 (c). This actually corresponds to the prediction of a thicker $\bar{\Theta}$, as expected. This observation is confirmed by the analysis of results in figure 15 for $x = 4\Lambda$ and $x = 16\Lambda$. Here, the average streamwise velocity profiles $\overline{u_x}$ are shown. Averages are performed in time and in the spanwise direction. These two sections have been chosen

as they are sufficiently upstream / downstream with respect of the first / last assimilation plane, respectively. The velocity profiles indicate that the estimator is closer on average to the DNS results, even far from the center-line where observation is available for $x = 6, 10, 14 \Lambda$ (gray dashed vertical lines). This trend is observed for $x = 4 \Lambda$ and for $x = 16 \Lambda$, downstream with respect to the last assimilation window. It appears as well that the convergence of the averaged results for the estimator is slower if compared with the pure LES case, but it is more in line with DNS. Thus, the present results seem to indicate that consistent average features of turbulent flows can be obtained via this sequential estimator, even including a limited amount of observation data.

These conclusions are reinforced by the observation of the isocontours of the time averaged normal velocity $\overline{u_y}$, which are shown in figure 16. Here, data are sampled at the streamwise section $x = 16 \Lambda$, downstream the very last observation window. The results indicate that the estimator produces results which are qualitatively in agreement with the DNS reference, as the same structure of the flow is conserved even with lower spatial resolution. On the other hand, LES data appear to be more diffused and the organization of the coherent structures is not the same.

The streamwise evolution of the momentum thickness $\overline{\Theta}$ is now investigated. The results, which are reported in figure 17, show that the LES overestimates the evolution of $\overline{\Theta}$. However, the distance from the DNS data is almost constant, showing that the maximum error produced by the Smagorinsky model comes for the prediction of the laminar state close to the inlet. On the other hand, the prediction produced by the sequential estimator is roughly in the middle between the LES and DNS results. This result comes with no operation on the model i.e. the Smagorinsky coefficient has not been modified. However, the augmented prediction could be actively used in a optimization procedure, in order to improve the performance of the LES model. While these analyses are not subject of investigation of the present research work, future development and applications will be discussed in section 7.

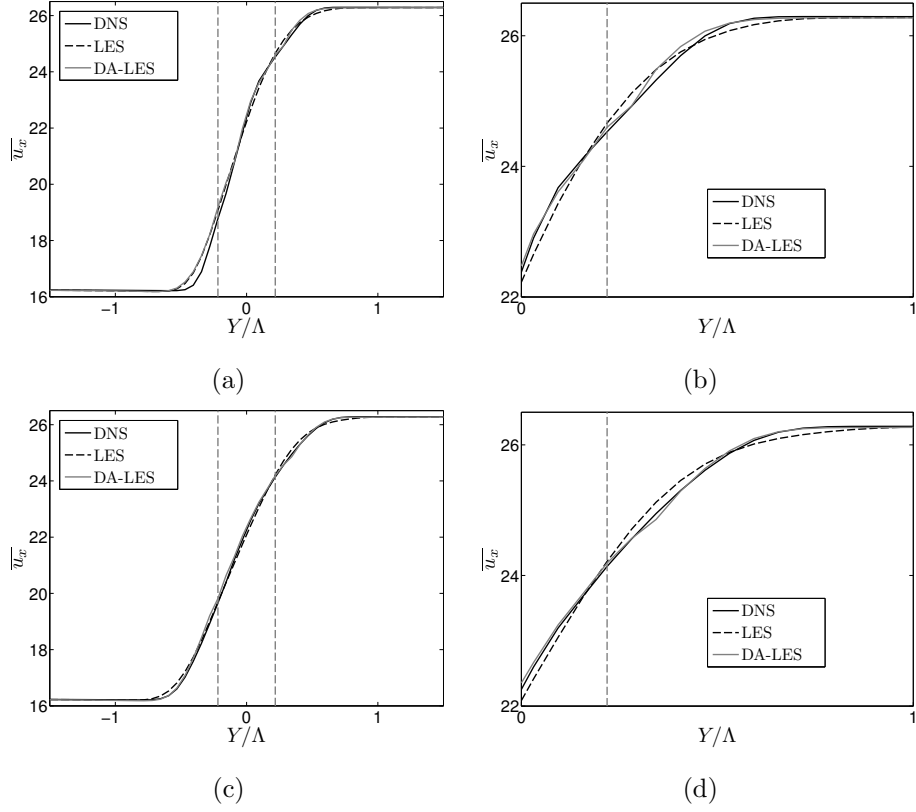


Figure 15: Averaged streamwise velocity $\overline{u_x}$ profiles. Averages are performed in the spanwise direction and over $T = 200 t_A$ times. The results are shown for (a-b) the streamwise section for $x = 4\Lambda$ i.e. in the transition region and for (c-d) the streamwise section for $x = 16\Lambda$ where the flow exhibits turbulent features.

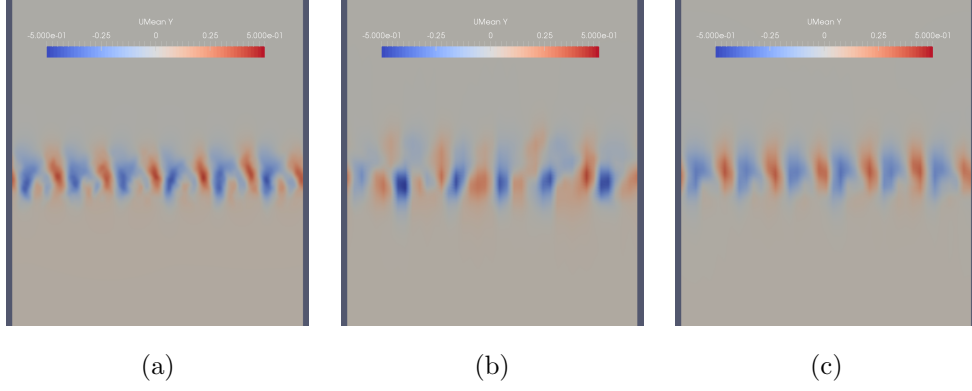


Figure 16: Isocontours of the time averaged normal velocity $\overline{u_y}$ taken at the streamwise section $x = 16\Lambda$. A zoom around the wake region is performed. Results for (a) the DNS calculation, (b) the LES simulation and (c) the *observer* estimator are shown, respectively.

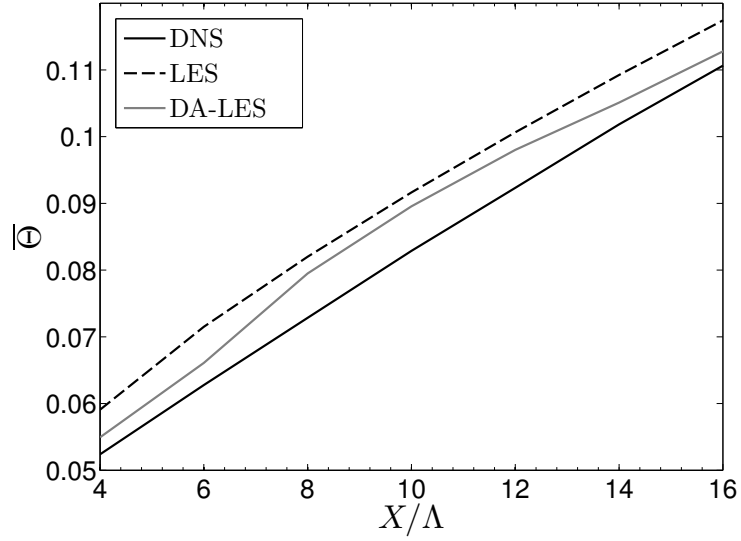


Figure 17: Streamwise evolution of the momentum thickness $\overline{\Theta}$. The results are reported for classical DNS and LES calculations, as well as for the sequential estimator.

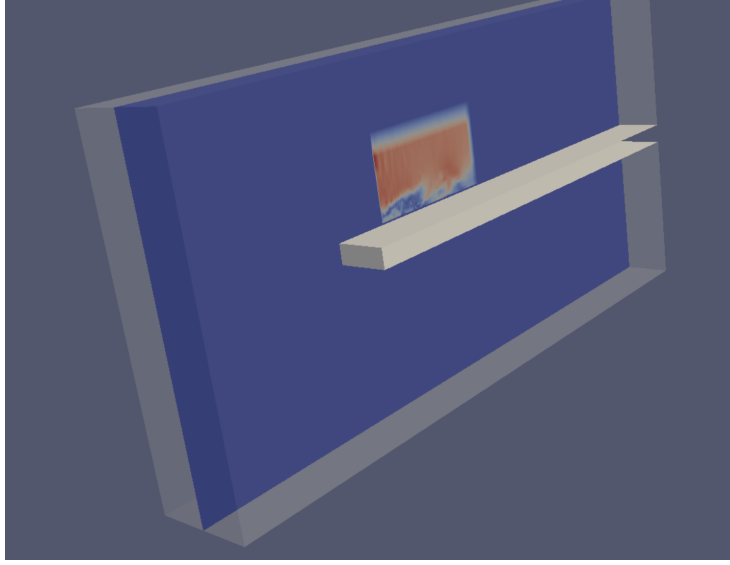


Figure 18: Representation of the test case of the flow around a thick plate. The plate profile is embedded in a three-dimensional physical domain used for DDES, while two-dimensional PIV data are provided for $z = 2.5H_p$.

6.2. Thick plate

The case of the flow around a thick plate of height H_p is investigated combining DDES numerical simulation [44] (*model*) and PIV experimental data [45] (*observation*). For the Reynolds number $Re = 80000$ here investigated, the flow exhibits full turbulent features. A recirculation bubble is observed at the trailing edge of the plate, whose average length is $\approx 5H_p$. This pioneering test case shows a high level of complexity, because of numerous problematic aspects:

- The numerical simulation domain is three-dimensional, while the PIV samples data is two-dimensional, as shown in figure 18. In addition, the observation window is rectangular and it covers a limited space in the physical domain investigated. Because of these two aspects, the transition between the assimilated state and the pure model prediction could be demanding on the numerical resolution of the pressure field, in order to smooth out this discontinuity.

- PIV data are sampled using an acquisition frequency of $4000Hz$, which is exactly $1/3$ of the characteristic advection time $t_A = H_p/U_\infty$. On the other hand, the time step imposed in the numerical simulation is around $5 \times 10^{-4}t_A$, which has been chosen in order to grant a maximum Courant number of $Co = 0.8$. This implies that the observation is available every ≈ 700 model time steps only.
- Preliminary comparison of average PIV results and DDES data show a difference of more than 10% in the mass flow rate, associated with different initial conditions / blockage ratio in the streamwise direction x (see figure 21). As a consequence, pure assimilation of the streamwise velocity could be problematic, leading to a badly-constrained optimization problem.

For this reason, a total of three numerical simulations has been performed:

1. A classical DDES simulation, which represents the *model*.
2. An estimator combining DDES and the normal velocity u_y of the PIV data, which will be referred to as DA-DDES1. The instantaneous flow field has been assimilated with a constant period of $t_A/3$. As previously mentioned, this corresponds to approximately 700 time steps of the numerical simulation.
3. A second estimator referred to as DA-DDES2. In this case, both u_x and u_y are assimilated. The interest here is to analyze the evolution of the velocity field, when a problematic constraint for the mass flow rate is imposed.

The physical domain investigated in the numerical simulations is $[-10, 25] \times [-8.5, 8.5] \times [0, 5]$ in H_p units. The origin of the system is located on the vertical surface of the thick plate. The physical domain is discretized in $\approx 7 \times 10^6$ elements. The recirculation bubble region has been refined in order to capture with sufficient resolution the wall dynamics. A preliminary simulation has been performed in order to dissipate the initial transient and provide a suitable initial condition for the three simulations performed. The *model* simulation has

been run for a total of $T = 100t_A$. Averages have been performed on the whole time length, which consisted of around 2.7×10^5 time steps. For the two estimators simulations, the total time length has been imposed to $T = 130t_A$. Data Assimilation begins from $t = 0$, while averages of the physical quantities have been started from $t = 30t_A$. This strategy allows for a degree of synchronization between the numerical model and the observation, before averages are calculated. As in the case of the spatially evolving mixing layer, second order schemes have been employed for the spatial / time derivatives. The only difference has been the use of a Linear-upwind stabilized transport (LUST) scheme for the divergence of the velocity field, which incorporates a second-order upwind scheme correction in the second-order centered classical scheme. The PIV data used in the present analysis are distributed over a uniform grid $x - y$ of 141×33 elements, for a total of 4653 samples per observation. Because of the very different structure of the experimental sampling grid with respect to the numerical mesh, PIV data have been interpolated on the center of the mesh elements in the observation window for $z = 2.5 H_p$, which is represented by the rectangle $x \in [3, 10]$, $y \in [0.6, 3.5]$ in H_p units. The horizontal surface of the thick plate is located at $y = \pm 0.5$, so that Data Assimilation at the wall is not performed. In order to simplify the analysis, the PIV interpolation error on the numerical mesh has been considered negligible with respect to the noise of the experimental measurement.

The matrices setting the level of confidence in the model / observation are now discussed. The error covariance matrix is set so that $P = 0.1 I$. This high level of uncertainty has been chosen in order to speed up the synchronization between the model and the observation. Similarly to the case of the spatially evolving mixing layer, the diagonal coefficients Q_i of the matrix Q have been locally determined as:

$$Q_i = C_{TP} \left(1 + \frac{\nu_T}{\nu} \right) \sqrt[3]{\Delta V^2} \Delta t \quad (30)$$

A visualization of an instantaneous field for the scalar field Q is shown in

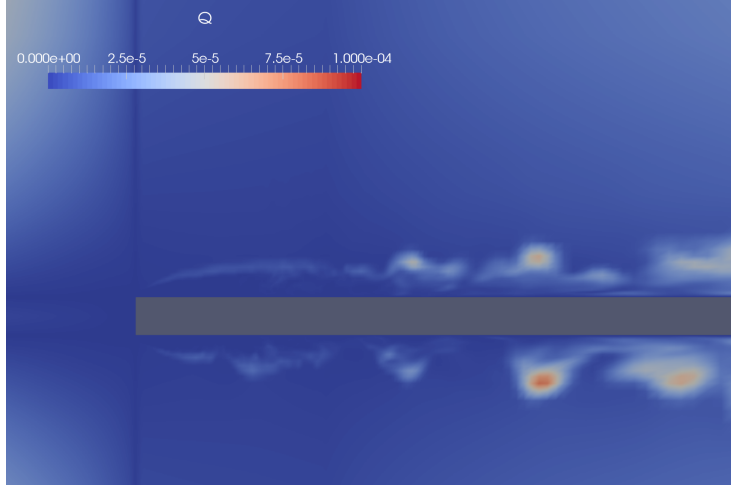


Figure 19: Visualization of the instantaneous diagonal elements ($t = 0$) for the matrix Q on the assimilation plane $z = 2.5H_p$.

figure 19. The value of the constant C_{TP} has been chosen in order to have roughly the same confidence in the model and in the observation during the update phase, considering that approximately 700 prediction steps are performed between each update. The precision of the PIV data is estimated to be of the order of 5%, so the covariance matrix R is fixed so that $R = 0.05^2 I$. In addition, in order to prevent problems close to the border of the rectangular assimilation window, a mask function M has been used to smooth out the value of the Kalman gain close to the borders. In this scenario, a sub-optimal gain in the form $\bar{K} = M K$ has been used to improve the stability of the results. The mask is defined as:

$$M = \exp \left(-250 \left(\frac{x - x_c}{x_c} \right)^8 - 25 \left(\frac{y - y_c}{y_c} \right)^8 \right) \quad (31)$$

where $x_c = 6.5$ and $y_c = 2.05$ are the coordinates of the center of the observation region. A representation of the mask function within the physical domain is shown in figure 20 (a), while a visualization of the diagonal component of the resulting Kalman gain for the streamwise velocity is shown in figure 20 (b).

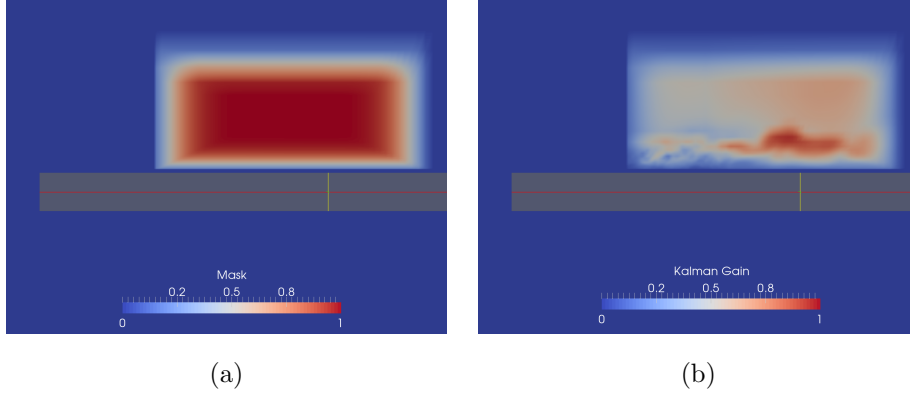


Figure 20: (a) Mask function M used to smooth out the Kalman gain when approaching the border of the observation region. (b) Instantaneous contours of the diagonal components of the Kalman gain for the streamwise velocity u_x on the assimilation plane, shown for $t = 100t_A$.

First of all, experimental and numerical data provide a very similar estimation of the recirculation bubble length, which is $\approx 5 H_p$. The two simulations performed using the estimator are now compared to DDES and PIV data. The case of Data Assimilation for the velocity u_y only (DA-DDES1) is discussed first. Average velocity profiles for $\overline{u_x}$ and $\overline{u_y}$ are reported in figure 21 for the streamwise sections $x = 5 H_p$ and $x = 6 H_p$. These sections are close to the middle of the observation window. The averages are performed in time and in the spanwise direction for the numerical simulation, while PIV data are clearly averaged over time only. The analysis of the normal velocity $\overline{u_y}$ in figure 21 (a) and (b) shows that the estimator successfully obtains a hybrid flow prediction for this physical quantity. The result can be closer to the pure DDES velocity field or to the PIV data depending on the streamwise section investigated. However, the analysis of spanwise averaged profiles for the numerical simulations indicates that the information is propagated by the Poisson equation in the physical domain, balancing the limited transmission of observed data via the *observer*. Further confirmation is obtained analyzing the L2 norm of the difference between time averaged numerical prediction for $\overline{u_y}$ and PIV data over

L2 norm error	DDES	DA-DDES1	DA-DDES2
N_{FP}	1.473	1	1.097

Table 4: L2 norm of the difference of the numerical profiles when compared to PIV data, for the average normal velocity $\overline{u_y}$. The norm has been calculated in the observation window for $z = 2.5 H_p$ and re-normalized over the value calculated for DA-DDES1.

the observation window:

$$N_{FP} = \int_{OW} \frac{\sqrt{(\overline{u_y} - \overline{u_{PIV}})^2}}{\sqrt{(\overline{u_{PIV}})^2 + 10^{-15}}} dS \quad (32)$$

The results, which have been re-normalized over the calculation for DA-DDES1, are reported in table 4. It appears that the average of the differences from PIV data and the *observer* prediction is approximately half with respect to the same norm for pure DDES (1.473), reinforcing the qualitative result of a flow configuration obtained imposing similar level of confidence in the model and in the observation. In figure 21 (c) and (d) averages velocity profiles for the streamwise velocity $\overline{u_x}$ are reported. PIV data are shown as well. In particular, the dotted line corresponds to a re-normalization of the experimental samples targeting the same average mass flow rate of the numerical simulation. Even if this quantity is not assimilated, the analysis of figure 21 shows that the prediction provided via the *observer* moves towards the shape of the normalized PIV data. This observation is fundamental, as it indicates that the whole numerical simulation process tends towards a high-confidence state even if a limited amount of samples is provided. Thus, assimilation of a physical quantity sampled in a small observation window can actually be beneficial for the prediction of other physical quantities in the whole domain investigated. This includes quantities that are difficult to measure via experimentation.

The analysis of the velocity profile for the second estimator DA-DDES2 (u_x and u_y assimilated) is now performed. Average velocity profiles at the streamwise section $x = 5 H_p$ are shown in figure 22. It is here reminded that, in the case of the numerical simulations, averages are performed both in time and in the spanwise direction z . The estimator provides an hybrid flow prediction from the

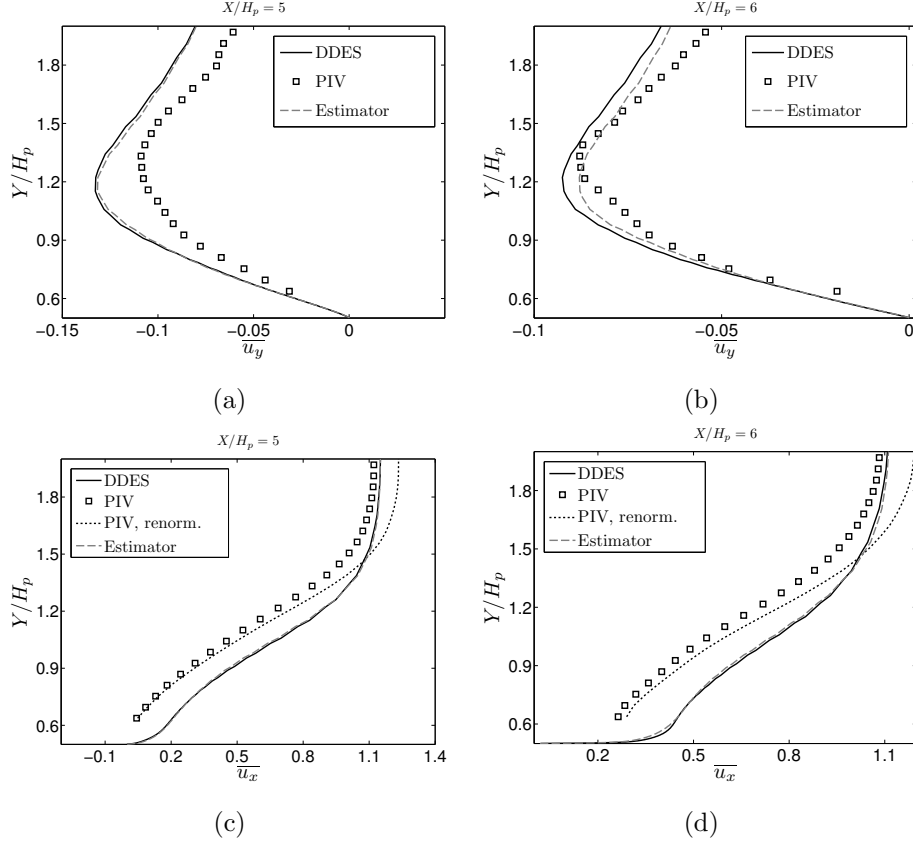


Figure 21: Averaged velocity \bar{u} profiles at the streamwise section (left column) $x = 5 H_p$ and (right column) $x = 6 H_p$. DDES results and PIV samples are compared with the flow prediction by DA-DDES1 estimator. The (first row) normal velocity \bar{u}_y and the (second row) streamwise velocity \bar{u}_x are reported. Averages are performed in the spanwise direction and over $T = 100 t_A$ times for the numerical simulations.

model / PIV data in the observation region, as shown for the averaged stream-wise velocity $\overline{u_x}$ in figure 22 (a). However, an augmentation of the predicted velocity by the estimator is observable close to the wall. This effect is the result of the over-constraint condition applied to the problem, as the estimator must (i) provide a weighted prediction based on model results / observation and (ii) to conserve the same mass flow rate of the model. The difference between the PIV data and the numerical simulations is measured on the observation window by the use of the quadratic function 32 for both $\overline{u_x}$ and $\overline{u_y}$. The analysis of the results for the DDES simulation and the DA-DDES2 calculation show that, on average, the estimator is significantly closer to the experimental prediction for $\overline{u_x}$ and $\overline{u_y}$ over the observation window. However, as the DDES and DA-DDES2 simulations must have the same inlet mass flow, the second must see and acceleration of the flow in the spanwise and normal directions, outside of the observation region. These results show how easy this numerical problem can be over-constrained, if the assimilation process is not structured carefully. In addition, present results stress as well the idea that suggest that the augmented flow must be actively used to optimize the parametric description of the model, in order to obtain a significant improvement in the overall prediction. In this case, the augmented prediction should be used to tune the value of the mass inflow rate at the inlet, in order to smooth out the differences between the classical DDES and the PIV observation. This topic will be discussed extensively in Section 7. The average prediction of the normal velocity is shown in figure 22 (b). In this case, the DA-DDES2 prediction is successful, as the averaged velocity profile correctly shows intermediate characteristics between the model and the observation. This is consistent with the rough hypothesis of similar level of confidence in the model and the observation that was initially set, and it is in line with results obtained by the DA-DDES1 estimator.

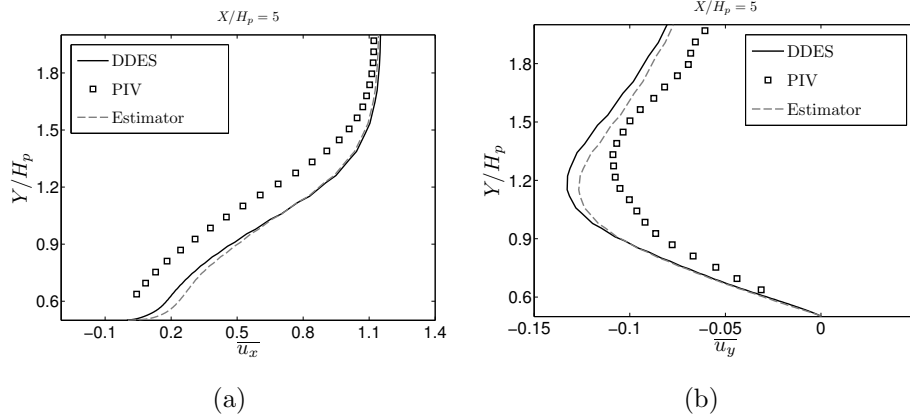


Figure 22: Averaged velocity \overline{u} profiles at the streamwise section $x = 5H_p$. Averages are performed in the spanwise direction and over $T = 100t_A$ times. DDES results and PIV samples are compared with the flow prediction by DA-DDES2 estimator. The results are shown for (a) the streamwise velocity \overline{u}_x and (b) the normal velocity \overline{u}_y , respectively.

7. Future works

The results presented in section 6 indicate that the augmented flow prediction performed via *observer* estimator successfully represents a dynamic system which accounts for the level of confidence in the model and in the available observation. To the Authors' knowledge, these analysis are among the very first successful Data Assimilation applications for the analysis of three-dimensional, unsteady turbulent flows. In particular, the statistics of several turbulence quantities have been investigated, which represent the main target objective in industrial applications. These Data Assimilation techniques have the potential to produce a breakthrough in the analysis of turbulent flows, in particular when model and observation are affected by completely different epistemic uncertainties and the level of confidence in the two basic investigative tools is roughly the same. However, the analysis of the two very different test cases in Section 6 highlights how results could see groundbreaking improvement accounting for the augmented information:

1. To improve *LES* subgrid scale modeling, in the case of the spatially evolving mixing layer. The possibility to locally optimize the parameters of the

subgrid scale model (i.e. model coefficient) would improve the results of the underlying LES simulation, reducing even more the difference observed when comparing the model and the estimator with DNS results.

2. To modify physical parameters of the DDES simulation, such as the mass flow rate, in the case of the the flow around the thick plate. This would allow the estimator to adjust the mass flow rate of the inlet, in order to avoid the over-constrained problem presented in section 6 for the estimator DA-DDES2.

Accounting for the augmented prediction in order to synergically improve the performance of the model is usually referred to as Linear Quadratic Gaussian (LQG) scheme in the framework of optimal control performed by sequential methods [46, 26, 47]. In summary, while present results show that sequential Data Assimilation of turbulent flows is achieved without exponentially increasing computational costs, they also open up to exciting perspectives for the optimization of numerical solvers under a new physical perspective. This last aspect represents a breakthrough in particular with respect to RANS modeling, where most of the well known models have been extended subject of analysis and tuning in order to improve the overall performances.

8. Conclusion

The present research work describes the integration of a Kalman filter based sequential estimator in the PISO algorithm of a segregated solver for incompressible flows. This technique, which is reminiscent of the proposal for coupled flow solvers by Suzuki [22], allows for the derivation of an *augmented* flow state accounting for the level of confidence in the model and in the observation provided. In addition, the present formulation naturally complies with a zero-divergence condition for the augmented state. Starting from this proposal, an extended analysis of the error covariance matrix P and the model / observation covariance matrices Q and R has been proposed. Because of the prohibitive

costs associated with a full resolution of these matrices, two reduction strategies have been proposed and assessed. These strategies dramatically reduce the increase in computational costs of the model, which can be quantified in an augmentation of 10% – 15% with respect to the classical numerical simulation. In addition, an extended analysis of the behavior of the elements of the matrix Q has been performed via analysis of simplified numerical cases. The results have shown that optimized values are linked to the truncation error of the discretization procedure. Simplified formulae accounting for the precision of the numerical schemes, which are set by the model user, have been elaborated and assessed.

The estimator has been applied to the analysis of a number of flow configurations exhibiting increasing complexity. The first case analyzed in the two-dimensional case of the flow around a cylinder, $Re = 100$. In order to test the capabilities of the model to synchronize model and observation, data from a previous numerical simulation have been integrated in the estimator in phase opposition. In addition, the observation window consisted of just 24 elements at a distance of $\approx 7 D$ from the center of the cylinder. The estimator proved to be very efficient in synchronizing the numerical model with the observed data. Similar results have been observed within the context of a three-dimensional simulation. In that case, the estimator produced the synchronization through a three-dimensional transient before returning to a pure two-dimensional evolution.

Finally, the observer has been used to investigate the statistical behavior of two turbulent flow configurations, namely the spatially evolving mixing layer and the flow around a thick plate. In both cases, the observation has been provided on a limited amount of mesh cells on a two-dimensional window, while the model resolved a full three-dimensional simulation. The analysis of the results has proven that in both cases the resulting augmented prediction correctly accounts for the level of confidence provided in the model and in the observation. However, these analysis indicated as well that these techniques, which are at an embryonic stage, can be extended using the *augmented* prediction as a powerful

tool for the optimization of the free parameters included in the numerical set-up.

The research work has been developed using computational resources within the framework of the project gen7590 DARI-GENCI. The Authors would like to thank F. Gava for the precious help provided in early stages of the work and Prof. E. Lamballais for the time he dedicated to fruitful discussion as well as for suggesting the use of the term *augmented* prediction.

9. References

References

- [1] J. Lumley, Stochastic tools in turbulence, Academic Press, 1970.
- [2] S. B. Pope, Turbulent Flows, Cambridge University Press, 2000.
- [3] D. C. Wilcox, Turbulence Modeling for CFD, D C W Industries, 2006.
- [4] P. Sagaut, Large-eddy simulation for incompressible flows. An introduction, third edition, Springer-Verlag, 2005.
- [5] P. A. Davidson, Turbulence. An introduction for scientists and engineers, Oxford University Press, 2004.
- [6] D. Simon, Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches, Wiley, 2006.
- [7] G. Artana, A. Camilleri, J. Charlier, E. Memin, Strong and weak constraint variational assimilations for reduced order fluid flow modeling, Journal of Computational Physics 231 (2014) 3264–3288.
- [8] R. Leroux, L. Chatellier, L. David, Bayesian inference applied to spatio-temporal reconstruction of flows around a NACA0012 airfoil, Experiments in Fluids 55 (2014) 1699.
- [9] S. B. Daley, Atmospheric Data Analysis, Cambridge University Press, 1991.

- [10] D. P. G. Foures, N. Dovetta, D. Sipp, P. J. Schmid, A data-assimilation method for Reynolds-averaged Navier-Stokes-driven mean flow reconstruction, *Journal of Fluid Mechanics* 759 (2014) 404–431.
- [11] V. Mons, J. C. Chassaing, T. Gomez, P. Sagaut, Reconstruction of unsteady viscous flows using data assimilation schemes, *Journal of Computational Physics* 316 (2016) 255–280.
- [12] C. Othmer, Adjoint methods for car aerodynamics, *Journal of Mathematics in Industry* 4 (2014) 6.
- [13] A. Onder, J. Meyers, Optimal control of a transitional jet using a continuous adjoint method, *Computers & Fluids* 126 (2016) 12–24.
- [14] Z. Sirkes, E. Tziperman, Finite Difference of Adjoint or Adjoint of Finite Difference?, *Monthly Weather Review* 125 (1997) 3373–3378.
- [15] R. E. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* 82 (1960) 35–45.
- [16] G. Evensen, The ensemble Kalman Filter for combined state and parameter estimation - Monte Carlo techniques for data assimilation in large systems, *IEEE Control Systems* 29 (2009) 83–104.
- [17] M. Chevalier, J. Hoepffner, T. R. Bewley, D. S. Henningson, A data-assimilation method for Reynolds-averaged Navier-Stokes-driven mean flow reconstruction, *Journal of Fluid Mechanics* 552 (2006) 404–431.
- [18] T. Suzuki, H. Ji, F. Yamamoto, Instability waves in a low-reynolds-number planar jet investigated with hybrid simulation combining particle tracking velocimetry and direct numerical simulation, *Journal of Fluid Mechanics* 655 (2010) 344–379.
- [19] M. C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, A. Trounev, Towards predictive data-driven simulations of wildfire spread - Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model

- for parameter estimation, *Natural Hazards and Earth System Sciences* 14 (2015) 2951–2973.
- [20] T. Suzuki, F. Yamamoto, Hierarchy of hybrid unsteady-flow simulations integrating time-resolved PTV with DNS and their data-assimilation capabilities, *Fluid Dynamics Research* 47 (2015) 051407.
 - [21] H. Kato, A. Yoshizawa, G. Ueno, S. Obayashi, A data assimilation methodology for reconstructing turbulent flows around aircraft, *Journal of Computational Physics* 283 (2015) 559–581.
 - [22] T. Suzuki, Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation, *Journal of Fluid Mechanics* 709 (2012) 249 – 288.
 - [23] R. I. Issa, Solution of the implicitly discretized fluid flow equations by operator-splitting, *Journal of Computational physics* 62 (1986) 40–65.
 - [24] H. Jasak, Error Analysis and Estimation for the Finite Volume Method with Applications to Fluid Flows The prediction and understanding of jet noise, Ph.D. thesis, Imperial Collge (1996).
 - [25] J. H. Ferziger, M. Peric, *Computational Methods for Fluid Dynamics*, Springer, 2002.
 - [26] J. Kim, T. R. Bewley, A linear systems approach to flow control, *Annual Review of Fluid Mechanics* 39 (2007) 383–417.
 - [27] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1960.
 - [28] D. G. Luenberger, Observing state of linear system, *IEEE Transactions on Military Electronics* 8 (1964) 74–80.
 - [29] M. Uhlmann, An immersed boundary method with direct forcing for the simulation of particulate flows, *Journal of Computational Physics* 209 (2) (2005) 448 – 476.

- [30] A. Pinelli, I. Naqavi, U. Piomelli, J. Favier, Immersed-boundary methods for general finite-difference and finite-volume Navier–Stokes solvers , *Journal of Computational Physics* 229 (24) (2010) 9073 – 9091.
- [31] D. Barkley, R. D. Henderson, Three-dimensional floquet stability analysis of the wake of a circular cylinder, *Journal of Fluid Mechanics* 322 (1996) 215–241.
- [32] H. M. Blackburn, R. D. Henderson, A study of two-dimensional flow past an oscillating cylinder, *Journal of Fluid Mechanics* 385 (1999) 255–286.
- [33] R. D. Henderson, Details of the drag curve near the onset of vortex shedding, *Physics of Fluids* 7 (1995) 2102–2104.
- [34] C. H. K. Williamson, Vortex dynamics in the cylinder wake, *Annual review of Fluid Mechanics* 28 (1996) 477–539.
- [35] C. Norberg, *Pressure Forces on a Circular Cylinder in Cross Flow*, Springer Berlin Heidelberg, 1993, pp. 275–278.
- [36] S. Ghosal, An analysis of numerical errors in large-eddy simulations of turbulence, *Journal of Computational Physics* 125 (1996) 187–206.
- [37] T. Colonius, S. K. Lele, Sound generation in a mixing layer, *Journal of Fluid Mechanics* 330 (1997) 375–409.
- [38] Y. Wang, M. Tanahashi, T. Miyauchi, Coherent fine scale eddies in turbulence transition of spatially-developing mixing layer, *International Journal of Heat and Fluid Flow* 28 (2007) 1280–1290.
- [39] W. A. McMullan, S. Gao, C. M. Coats, The effect of inflow conditions on the transition to turbulence in large eddy simulations of spatially developing mixing layers, *International Journal of Heat and Fluid Flow* 30 (2009) 1054–1066.

- [40] M. Meldi, M. V. Salvetti, P. Sagaut, Quantification of errors in large-eddy simulations of a spatially evolving mixing layer using polynomial chaos, *Physics of Fluids* 24 (2012) 035101.
- [41] J. Smagorinsky, General circulation experiments with the primitive equations. I: The basic experiment, *Monthly Weather Review* 91 (3) (1963) 99–165.
- [42] D. K. Lilly, The representation of small-scale turbulence in numerical simulation experiments, *Proceedings of the IBM Scientific Computing Symposium on Environmental Sciences*, 1967, pp. 195–210.
- [43] M. Meldi, D. Lucor, P. Sagaut, Is the Smagorinsky coefficient sensitive to uncertainty in the form of the energy spectrum?, *Physics of Fluids* 23 (2011) 125109.
- [44] M. L. Shur, P. R. Spalart, M. K. Strelets, A. K. Travin, A hybrid RANS-LES approach with delayed-DES and wall-modelled LES capabilities, *International Journal of Heat and Fluid Flow* 29 (2008) 1638–1649.
- [45] C. Sicot, R. Perrin, T. T. Tran, J. Boree, Wall pressure and conditional flow structures downstream of a reattaching flow region, *International Journal of Heat and Fluid Flow* 35 (2012) 119–129.
- [46] T. R. Bewley, Flow control: new challenges for a new Renaissance, *Progress in Aerospace Sciences* 37 (2001) 21–58.
- [47] S. L. Brunton, B. R. Noack, Closed-Loop Turbulence Control: Progress and Challenges, *Applied Mechanics Review* 67 (2015) 050801.